Lecture Notes in Economics and Mathematical Systems 519

Springer-Verlag Berlin Heidelberg GmbH

Andreas Klose
M. Gracia Speranza
Luk N. Van Wassenhove
(Eds.)

Quantitative Approaches to Distribution Logistics and Supply Chain Management

Springer

Editors

Priv.-Doz. Dr. Andreas Klose
University of St. Gallen
Bodanstr. 6
9000 St. Gallen, Switzerland

Prof. Dr. Luk N. Van Wassenhove
INSEAD
77305 Fontainebleau Cedex, France

Prof. Dr. M. Gracia Speranza
University of Brescia
Department of Quantitative Methods
C. da S. Chiara, 48B
25122 Brescia, Italy

Library of Congress Cataloging-in-Publication Data

Klose, Andreas, 1963-
Quantitative approaches to distribution logistics and supply chain management / Andreas Klose, Grazia Speranze, Luk N. Van Wassenhove.
p. cm.
Includes bibliographical references.
ISBN 978-3-540-43690-4 ISBN 978-3-642-56183-2 (eBook)
DOI 10.1007/978-3-642-56183-2
1. Business logistics. 2. Materials management. I. Speranza, M. Grazia, 1957- II. Wassenhove, L. N. van (Luk N.) III. Title.

HD38.5 .K586 2002
658.7--dc21

2002070481

ISSN 0075-8450
ISBN 978-3-540-43690-4

http://www.springer.de

Typesetting: Camera ready by author
Cover design: *Erich Kirchner*, Heidelberg

Printed on acid-free paper 55/3111 5 4 3 2 1

Editorial

Logistics management is concerned with the design and control of efficient and cost-effective flows of material and information through complex networks from point of origin to point of consumption. Increased international competition and an increased need of quickly confirming to customer requirements despite longer distances for distribution and a growing product variety stresses the importance of distribution logistics, that part of logistics management responsible for delivering products to customers at the right place at the right time in the right condition for the right cost. Physical distribution is just a part of the supply chain. Effective distribution management is, however, impossible without taking the strong links to procurement and production as well as the interrelations between other logistic processes and parties involved in the supply chain into account. Therefore, there is no clear-cut dividing line between logistics in general, distribution logistics and supply chain management.

Interest in logistics and supply chain management, both in industry and in academia, has grown rapidly over the past years. On the one hand, this trend is due to the enormous potentials in improving logistics efficiency exploitable by means of intelligent planning techniques and improved coordination of logistic processes. On the other hand, this trend is caused by the development of information and communication systems that are able to provide access to comprehensive data from all components of the supply chain. Vendors of supply chain management software are going to add "business intelligence" components to their systems, which not only allow to access, share and transfer data in the supply chain but also to utilize this information in order to improve decision making with the help of decision-support systems. Such systems heavily rely on quantitative models and techniques developed in the field of Operations Research in the last decades. Today, there is no doubt of the importance of quantitative techniques in our modern business environment. Fortunately, advances in quantitative models and methods and in their applicability to practical logistic problems are still achieved. Trends like e-commerce, the globalization of markets and the need of integrating reverse flows in the supply chain add to the growing complexity of logistic networks and require even more effective models and algorithmic tools. The papers collected in this book contribute to some of these new developments in quantitative approaches to distribution logistics and supply chain management.

The main orientation of the book is not towards the theory underlying the employed methods but towards practical problem solving.

The volume in hand continues a series of books, which are the outcome of the work of a group of researchers who have met at a number of "International Workshops on Distribution Logistics (IWDL)" since 1994. This book includes reviewed papers that were presented and discussed during IWDL 5, at Fontainebleau (France) in October 1999, and during IWDL 6 at St. Gallen (Switzerland) in February 2001.

We have organized the 22 papers in seven Chapters. The first three chapters address general issues in supply chain management, in the relatively new field of reverse logistics as well as new challenges to distribution logistics caused by the evolution of e-commerce. The other four chapters deal with main functions of distribution logistics: strategic and tactical planning of distribution networks; operational, tactical as well as strategic problems related to vehicle routing and transportation; tactical and operational issues internal to the production center or the warehouse; and finally inventory problems.

Chapter 1 is concerned with various important topics in the field of supply chain management. The paper by *Vis and Roodbergen* first introduces basic supply chain concepts and afterwards analyses the impact of various trends on supply chain performance. The importance of cycle time reductions, the influence of reverse logistics, e-commerce, third party logistics and global logistics and the resulting threats and opportunities are analysed by means of a number of case studies. In his contribution, *Blackburn* outlines a methodology for valuing response time in supply chains. Using results of inventory theory, he establishes important properties of the marginal value of time, which are at first sight astonishing: Firstly, the marginal value of time increases with decreasing response time. Secondly, for equal response times, the marginal value of time is greater at non-optimal inventory levels than at the optimum. The paper of *Fjell and Jørnsten* concludes Chapter 1. They study the important question, how coordination between supply chain partners can be achieved by means of pricing mechanisms, which result in a locally rational behaviour that is also efficient from a global perspective. They propose a novel negotiated two-part tariff scheme and argue that this pricing mechanism is a good means to achieve channel coordination.

Chapter 2 exclusively addresses managerial problems and solution methods in the field of reverse logistics. Over the past years, environmental problems have reinforced public interest in reuse and recycling. Take-back and recovery of used products leads to additional goods flows from the user back to the producer. Reverse logistics is concerned with the management of these opposite flows. Since product recovery affects product design, procurement, production and forward distribution, the challenge is to integrate forward and return flows and to obtain integral closed-loop supply chains. *Guide and Van Wassenhove* give an overview of the field of closed-loop supply chains. They summarize various cases of reused products, elaborate the differences between

branches and elicit main managerial problems and success factors. Furthermore, important research issues in remanufacturing are pointed out. The paper by *Krikke, Pappis, Tsoulfas and Bloemhof-Ruwaard* gives an overview of design principles of reverse logistics, which extend the scope and applicability of design rules for forward supply chains to reverse supply chains. The presented principles are applied in a case study and may provide a checklist for improving closed-loop logistic systems. *Mazzarino, Pesenti and Ukovich* consider logistic system optimization for a reverse logistic case. In contrast to traditional approaches, they propose, however, a multi-agent approach, which takes actors' behaviour and multiple decision makers into account. The paper by *Gotzel and Inderfurth* examines an extended MRP approach, material requirements and recovery planning (MRRP), for production control in a system with external stochastic return flows and stochastic demand. They show that the application of MRRP leads to near-optimal results. *Bloemhof, van Nunen, Vroom, van der Linden and Kraal* describe a practical reverse logistic problem that emerged at a dairy producer in the The Netherlands. A cost evaluation tool based on scenario analysis for selecting between different packaging systems is developed. In addition to traditional costs, environmental costs and aspects are also taken into account.

Chapter 3 addresses the influence of e-commerce on distribution logistics. The papers in this chapter describe threads, opportunities and new problems that have to be solved in order to cope with the challenges resulting from the evolution of e-commerce. The paper by *de Koster* is concerned with the question of how to organize logistic fulfilment processes in a BtC e-commerce environment. Different alternatives for designing effective distribution structures are pointed out and a model relating a company's objectives to characteristics and choices in distribution is presented. The need of quickly responding to diverse customer needs in BtC e-commerce raises the complexity of delivery processes and timely delivery gets more difficult. *Daduna* discusses this problem and proposes a heuristic method for solving routing problems with tight time windows that arise in electronic retail trade. The spread of the Internet has also significantly increased the use of Internet auctions for exchanging goods among and between individuals and companies. *Bjørndal and Jørnsten* consider combinatorial auctions where bidders bid on bundles of items and where the value of an object to a participant depends on what other objects the participant acquires. Bjørndal and Jørnsten employ sensitivity analysis and linear programming duality in order to solve the pricing problem and to derive a feedback mechanism providing information to bidders that may give the bidders incentive to change their bids.

Chapter 4 treats strategic planning of distribution networks as well as allocation problems arising in tactical network planning. The paper by *Bauer* discusses practical problems of data gathering, data validation and insufficient data quality in distribution system design and gives hints on possible solutions elaborated on two case studies. The paper by *Romeijn and Morales* considers the problem of assigning plants to warehouses and customers to

warehouses in a multi-level distribution network with time-varying demands. They present a mixed-integer model for the minimization of transportation, production and handling costs, and derive an effective greedy heuristic for solving the model. *Klose and Drexl* describe assignment type optimisation problems arising in logistic system analysis and propose a solution method based on problem partitioning and column generation.

The topic of Chapter 5 is transportation planning and vehicle routing. The paper by *Angelelli and Mansini* deals with a vehicle routing problem with time windows and simultaneous pick-up and delivery. A branch-and-price algorithm for computing optimal solutions is proposed. *Angelelli and Speranza* apply a vehicle routing model to estimate the operational costs of different waste collection systems and to support decision making regarding the type of system to adopt. Although vehicle routing is usually treated as an operational or tactical issue, the determination of efficient vehicle routes can be a strategic problem if stable routes are required. *Dillmann* summarizes the experiences made when solving a large number of strategic vehicle routing problems for press wholesalers. He addresses important practical problems related to data measurement, data validation and model building in the presence of soft constraints and multiple objectives. For solving large-scale vehicle routing problems a dialog-based procedure is proposed. Furthermore, it is shown how the implementation of computed routes in practice can be supported.

Chapter 6 is concerned with tactical and operational issues of warehousing. The paper by *Chevalier, Pochet and Talbot* presents analytical results from queuing theory for estimating the number of vehicles needed in an automated material handling system. The model can provide a methodology for designing automated guided vehicle systems and is validated by means of simulation. *De Koster and van der Meer* compare the performance of on-line and off-line rules for dispatching vehicles in internal transport systems. The authors show that for different layouts of the transportation system off-line optimisation attains high performance if the system is relatively idle; however, in high throughput environments the proposed on-line dispatching rules attain high performance.

Finally, Chapter 7 addresses topics in inventory control. *Laan and Teunter* compare average cost models with approaches based on the net present value for single-source, multi-source, multi-stage inventory systems and for a system with remanufacturing and disposal. The authors show in particular that for complex inventory systems there is a considerable performance gap between the widely used average cost and the net present value approach. The paper by *Wagner* treats the problem of determining safety stocks in capacitated single-stage multi-product production-inventory systems. Simple schemes for calculating safety stock levels are proposed and their reliability shown by means of simulation experiments. The paper by *Smits and de Kok* considers the impact of freight consolidation policies on the lead time, which influences inventory requirements. The authors derive approximations for the

lead time behaviour where items are consolidated according to different types of consolidation policies.

Unfortunately, a tragic event overshadowed the making of this book. In September 2001, one of the authors, Roland Dillmann, died unexpectedly and far too early in the age of 56 years. Roland Dillmann was professor for mathematical methods in Economics at the University of Wuppertal from 1975 until his death. He teached Economics, Statistics, Econometrics and Operations Research. Furthermore, he was engaged in the administration of the faculty and acted successfully for a long time as consultant for press wholesalers in fields like transportation planning and demand forecasting. We have not only lost an excellent researcher with an extremely broad knowledge but also a very good friend. We are grateful to Simon Görtz, University of Wuppertal, and Thomas Bieding, Dillmann&Co GmbH, for their help in reediting parts of Roland Dillmann's paper that is published in this volume.

Acknowledgement

The editors would like to thank the authors of the papers for their contribution. All papers submitted for publication in this volume have been subject to a refereeing process and we are grateful to the referees whose work was essential to ensure a high quality level of this book.

Last, but not least, the editors are deeply indebted to Prof. Dr. Paul Stähly. Paul Stähly was professor for Operations Research at the University of St. Gallen from 1973 until his retirement in March 2001. Since the first "International Workshop on Distribution Logistics" in 1994, Paul Stähly was an active and leading member of our "IWDL group" who invested his energy in strengthening the cooperation between the group members. Furthermore, he was main organizer of the workshop in St. Gallen. Without his support, this workshop and thereby this book would not have been possible. Therefore, we would like to dedicate this book to Paul Stähly as a recognition for his merits for our "IWDL group".

Dr. Andreas Klose, University of St. Gallen, Switzerland
Prof. Dr. M. Grazia Speranza, University of Brescia, Italy
Prof. Dr. Luk N. Van Wassenhove, INSEAD, Fontainebleau, France

May 2002

Contents

Chapter 1: Supply Chain Management

Chapter 2: Reverse Logistics

Chapter 3: Distribution Logistics and E-Commerce

Chapter 4: Warehouse Location and Network Planning

Chapter 5: Vehicle Routing and Transportation

Chapter 6: Warehousing

Chapter 7: Inventory Control

Appendix

Chapter 1

Supply Chain Management

Examining Supply Chains from Practice

Iris F.A. Vis and Kees Jan Roodbergen

Erasmus University Rotterdam, Rotterdam School of Management/Faculteit Bedrijfskunde, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Abstract. Companies have to adjust their actions constantly in order to remain competitive in the current world of Internet and e-commerce. In this paper we first give an introduction in supply chains concepts and trends that impact the performance of the supply chains. The issues we discuss include the importance of cycle times reductions, the influence of reverse logistics, e-commerce, concentration and centralisation, third party logistics and global logistics.
Some practical consequences, threats and opportunities are sketched for six supply chains from real life. For example, the supply chains of mobile phones, glass and design furniture are studied. The intention of this study is to shed some light on the current state of logistic innovations in Dutch companies by means of a number of case studies.

Keywords. Supply chain, supply chain management, performance improvement, cycle time reduction, case studies.

1 Introduction

Nowadays, products can be ordered electronically with computers or mobile phones via the Internet. As a consequence of such a fast ordering process, customers expect a fast delivery process as well. Existing suppliers have to adapt to this new situation to remain competitive. Furthermore, new organisations emerge to deal with this new challenge.

The *supply chain* encompasses all activities associated with the flow and transformation of goods from the raw materials stage through to the end-user, as well as the associated information flows (Handfield and Nichols (1999)). Historically, the focus of many companies has been on improving processes within the company. More recently the awareness has grown that it is also necessary to improve the process of interactions between firms to keep up with customer demand. Activities of multiple companies within a supply chain should be adjusted to ensure an efficient supply chain instead of an efficient company within a potentially inefficient supply chain. As a result, involvement in the entire supply chain is a necessity to be able to meet, for example, tight delivery schedules and high service levels. The efforts to harmonise the processes in a supply chain are called *supply chain management*. Co-operation, effective coordination of materials and informa-

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

tion and confidence through the entire supply chain are necessary to obtain a valuable chain with satisfied customers.

The costs of the flow of materials through the supply chain can approach to 75% of the total budget (Handfield and Nichols (1999)). This material flow should be managed accurately to ensure that the corresponding costs can be reduced, while meeting the service levels simultaneously. Therefore, attention should be paid to logistics activities, such as planning of transport, planning of activities within warehouses, determining inventory levels and determining locations of distribution centres.

Logistics can be defined, according to the *Council of Logistics Management*, as that part of the supply chain process that plans, implements, and controls the efficient, effective flow and storage of goods, services and related information from the point of origin to the point of consumption in order to meet customers requirements. This includes, for example, the transportation of goods from one location to another as well as inventory keeping in warehouses. Inventories are – among other reasons – held to balance fluctuations in production and demand, to combine products from several producers for delivery to common customers and to enable emergency deliveries of critical components. However, logistics is not just concerned with handling products. The information flows to ensure that the right product is at the right place at the right time in the right quantity are part of logistics as well.

As described, information is one of the flows through the supply chain. Customers just present their wishes to retailers. From the retailers this information flows to distribution centres and thereafter to the manufacturer. Next, the manufacturer can fulfil the wishes of the customers. In the past information was generally passed through to other members of the supply chain on paper. This was a slow, error prone and expensive process. However, for successful management of supply chains, well organised information flows are of great importance. Good information flows can, for example, reduce inventories because of reduced uncertainty. Furthermore, to serve customers in the best way it is important to have information on orders, product availability and so on available on the fly. Information should be available for various departments within one organisation and for organisations across the supply chain. Therefore, a paperless environment should be created. More details on this are given in section 2.2.

To measure the performance of a supply chain, the performance of an individual part, such as a warehouse, or the performance of the complete supply chain can be observed. Several performance measures can be defined for measuring the complete supply chain. Firstly, the total cycle time of the supply chain, which is the time from the stage of raw materials to the delivery of the product to the customer can be used as a performance measure. Secondly, the changes in average inventory through the supply chain and thirdly, the reliability of the quality of the product are potential performance measures.

In section 3, six actual supply chains are discussed. Success factors and suggestions for performance improvements of these supply chains are given. Firstly, a number of trends that impact the supply chain, are discussed in section 2. For almost all of the cases described in section 3 a corresponding trend can be found in section 2.

2 Trends

In this section we discuss a number of trends that potentially impact the supply chains. More trends and forecasts for the developments of these trends in supply chains in 2008 are given in Carter *et al.* (2000).

2.1 Cycle Time Reduction

The cycle time of a supply chain equals the total time required to complete the total process from raw materials to the delivery of the finished product to the customers. According to Handfield and Nichols (1999) only a small amount (5%) of the cycle time is used for executing real processes. The rest of the time is spent on activities like waiting. To remain competitive and to satisfy demands of customers the complete process should be executed effectively. Therefore, one of activities in supply chain management is to make improvements in cycle time performance. This increases the flexibility to react on changes in customer demand, it decreases the risk of unsaleable stock and reduces inventories.

To shorten the cycle time of the supply chain, we can focus on the separate processes in the supply chain. The key processes are processes with the longest average throughput times and processes with the greatest variance in throughput times. Possible candidates for review are warehouses, planning and scheduling of materials, transportation of goods and the manufacturing processes.

One radical way to shorten the cycle time is to design a supply network without warehouses. Products are then shipped directly or via cross-docking centres from the manufacturing centre to the customers. Cycle times are mainly decreased because products are no longer stored (i.e. waiting) in warehouses. However, feasibility of this option may depend on several factors, such as flexibility of the production, communication facilities between companies and reliability of production and distribution.

A less rigorous way is to reduce the number of warehouses within the supply network. Traditionally, inventory is held at several locations in the supply chain, such as at the retailer and at the distribution centre. A reduction in the number of inventory points is possible, for example, by concentrating inventory at the distribution centre. See section 3.6 for an example. As a result, all warehousing activities of the retailer become superfluous and savings in time and costs can be obtained.

Important in this respect is also the just-in-time concept; it does not matter how long it takes for a product to arrive as long as it arrives exactly at the time it is needed to satisfy demand of customers.

2.2 E-commerce

E-commerce is the collective noun for all techniques, which take care of the fact that all transactions can be executed in a paperless environment. EDI, Internet and

e-mail are examples of these kinds of techniques. Electronic Data Interchange (EDI) is one of the systems that can be used to enable faster and more reliable delivery of orders to warehouses and goods to the customers. With EDI standard documents are exchanged between computers. For example, orders of customers, requests for production and so on. With EDI information can be exchanged, in contrast to paper based systems, within a few minutes. As a result, an effective inventory management, a better service and a better way of communicating can be obtained. Besides EDI, the Internet is used more and more. In Rogers *et al.* (1996) it appeared that using these kinds of technologies in warehouses results in a significantly better performance in the areas of quality improvement, cycle time reduction and productivity improvement.
On the Internet suppliers can offer their products against fixed prices. Another way, to buy or sell products on the Internet is to participate in an auction. With Internet all aspects of the traditional purchase process can be executed. The customer can search on the Internet for information on the product and potential suppliers. The customers can buy the products of their choose via the Internet without actually having seen them. The payment can be done by filling out a credit card number. Next, the retailer sends the products directly to the customer. Sometimes, it might be possible for the customer to indicate the preferred time of delivery. Afterwards, the service continues via the Internet. Questions concerning the product can be asked by e-mail. The only physical flows are the delivery of the item and potentially some spare parts and return flows (see section 2.4). See also Hagdorn-Van der Meijden and Van Nunen (2000).
Customers order more frequently and in smaller amounts. Furthermore, they want their products to be delivered as fast as possible. Therefore, one of the most important consequences of Internet is the fact that the size of orders becomes smaller while the number of orders grow simultaneously. As a result, the order processing, the inventory management, the layout of the warehouse and so on, may have to be changed for efficiency. The delivery times should become smaller and more flexible. The supply chain is directed straight by customers. Furthermore, 'power relations' within the supply chain will shift. For example, from the retailer to the manufacturer if products are delivered directly to the customers by the manufacturer.

By using e-commerce, supply chains can become more competitive. For example, costs can be reduced and information can easily be distributed through the supply chain (see also Fraser *et al.* (2000)).

2.3 Concentration and Centralisation

Firms that expand their business and increase their production capacity often spread out their activities over a number of locations. There are basically two ways by which the number of locations of a firm can increase. Firstly, by building new locations, where a part of the activities or all types of activities are carried out. Secondly, by taking over other firms, for example small family businesses. As

a result of either option, the firm has a number of production locations through out the country or world. When expansion occurs by taking over other firms, then the management structure is generally not homogeneous. Each location may have its own goals, markets, computer systems and so on. To ensure an efficient supply chain, an option is to centralise the management function. Decisions concerning, for example, the product assortment, planning of transport between production locations and customers and planning of orders over various subsidiaries can be made centrally.

A prerequisite for centralisation of management is co-operation of all subsidiaries. This can be a problem if the right to make decisions has to be taken from firms that were taken over. Secondly, well organised and fast information flows between all locations of the firm should be established. As a result, the costs within the supply chain will be lowered and the supply chain will be more transparent for employees and customers.

Expanding the business by taking over other businesses instead of building new firms at chosen locations, can result in illogical choices for locations of new subsidiaries of the firm. As a result, the supply chain will function less efficient than possible. This problem also occurs if a firm has too many locations, where activities can be carried out. Concentration, i.e. uniting different parts of the firm, is a possibility to deal with this problem. A large reduction in costs might obtainable. For example, transport between some locations becomes a superfluous activity. By concentrating subsidiaries of a firm, centralisation of management can be obtained more easily.

2.4 Reverse Logistics

Return flows in the supply chain consist, for example, of products that customers ordered via the Internet but decided not to buy, products that are defective and need to be repaired and products that have reached the end of their life cycle and need to be dismantled, remanufactured or recycled. New products that return need to be checked before they can be sold again. These return flows are managed by the receiver. There is a lot of uncertainty in return flows. Reuse of products, components and materials, initiated by the desire for environmental improvements, is enforced by legislation. It may induce cost savings, but can also constitute a competitive advantage, due to the 'green image' of the company. Return flows are thus becoming increasingly important in logistics.

It is clear that several adaptations in the logistics branch are needed to respond to this trend. In Fleischmann (2001) some new logistics concepts for the handling of return flows are discussed.

2.5 Third Party Logistics

As described in the previous sections, customers are requiring more and more of their suppliers. As a result, high demands are made on production and distribution.

Each firm should decide whether all activities can be executed or that only key activities should be performed by the firm itself. The remaining activities can be outsourced to third parties. The relationship between firms and third party logistics providers features mutual trust, respect and openness. This is necessary to obtain strategic competitive advantages.

Warehousing and transport are two of the main logistics activities, that can be outsourced. In 1996, 50% of all transport activities in The Netherlands was executed by third party logistics providers. Furthermore, 15% of the warehousing activities were outsourced (see also Van der Baan (1997)).

Advantages of outsourcing side activities are, for example, reduction of costs, better service, larger flexibility and efficiency, less investments in logistics and skilled staff. On the other hand, the direct contact between the firm and the customer disappears. Furthermore, confidential information has to be shared with the logistics provider. This might be a disadvantage, if the provider also works for the competitor.

To ensure that the supply chain is managed efficiently, the outsourcing process should be well implemented. The third party logistics provider should function effectively as part of the complete supply chain.

2.6 Global Logistics

Supply chains are no longer concentrated within one country, but they are spreading out all over the world and flows of material are crossing borders. Supply chain managers nowadays have to deal with the uncertainty and complexity of global networks. Numerous differences exist between national and global supply networks. We will discuss a few of them (see also Dornier *et al.* (1998)).

Within global networks geographic distances will be larger and time differences will occur. Firms are dealing with this problem by keeping larger inventories. Delivery times will be larger. A larger variance in delivery times will also occur due to unexpected delays, like customs regulations. Firms face a great challenge in implementing just-in-time production while suppliers are at a great distance.

Secondly, forecasting demand in foreign countries is more difficult than in the home country. Firms are operating in environments with different languages, different cultures and habits. As a result large safety stocks are held. An other influence on the performance of the global supply chain is the effect of the change of exchange rates. They influence costs, prices and the amount of products sold.

If firms are starting operations abroad, they have to deal with problems concerning infrastructure. In, for example, developing countries there is a lack of sources, like transportation networks, telecommunication networks, skilled employees and materials. Furthermore, firms should be aware of (changes in) foreign regulations.

As a result, supply chain managers should be aware of local regulations, conditions, habits and so on when implementing a global network. Still, in global net-

works it is important to reduce inventories, to reduce the cycle time of supply chains and to pay attention to all other trends mentioned.

3 Practical Cases

Six case studies have been performed to investigate the current state of Dutch companies with regards to the trends we described in section 2. The goal of the case studies was three-fold. Firstly, the intention was to identify the trends that impact the various supply chains. For some case studies we also identified possible future developments. Secondly, it was investigated which actions were taken by the companies to react on these trends. Thirdly, additional actions were identified that companies could take to further improve the supply chain efficiency. The latter aspect was merely added to illustrate one of many possibilities a company has to adjust to the observed trends. It is by no means intended to be an exhaustive list of possible actions. In each case study several research methods have been used, such as observation, literature study and taking interviews. In the case study on return flows for food retailers a questionnaire was sent to a number of retailers.

3.1 Supply Chain of Glass

Glass is produced by using the raw materials sand, soda, lime and aluminium, magnesium or oxide. In a factory these raw materials are put together in a furnace for four days. Glass can be made in one size with various thicknesses. Changing to another thickness is a very time-consuming and expensive activity. The glass can be worked up to for example carglass, layered glass and sun-blind glass. Glass can be stored in warehouses. However, glass can taint and as a result it cannot be stored for more than one year. From the warehouses at the production site, glass is transported to other production sites or to warehouses of wholesalers. At the other production sites, glass is worked up to, for example, isolation glass. At the wholesaler the glass is cut into standard sizes and stored in a warehouse until orders of customers arrive.

The supply chain that has been studied, has experienced an enormous growth in the past 15 years. All parts of the supply chain belong to a single company. The different production stages of glass occur at various factories. As a result, frequent transport of glass is needed to produce the final product. This is a time-consuming and expensive (high probability of damage) process.

To expand its business, the company has taken over small wholesalers. The existing management and organisational structures of each small wholesaler have remained the same after a take-over. The various subsidiaries (23) are spread out over The Netherlands. However, some subsidiaries are quite close to each other. Most wholesalers still carry their own name instead of the name of the owner. Therefore, it is not clear to most customers that they purchase glass from a multi-

national. As a result, customers order their glass at a wholesaler they are familiar with, instead of at the nearest wholesaler.

There is no organised communication among wholesalers concerning transport or order handling. Each wholesaler has its own profit target, i.e. the wholesalers behave like competitors of each other. Thus, an order is not fulfilled from the wholesaler established closest to the customer, but by the wholesaler that received the order. Also, transportation is not coordinated between wholesalers, allowing multiple partially filled trucks travelling to nearby destinations. Clearly, transportation costs are higher than necessary.

To obtain an efficient and effective supply chain co-operation between all parts of the supply chain is needed to deal with the problems mentioned. When solving the problems it should be kept in mind that the service to customers has to remain equal or increase. Suggestions to improve the supply chain were found by observing the current situation and by interviewing people at wholesalers and factories.

First of all, it can be concluded that there are probably too many wholesalers. From literature it is known that inventory keeping at a single location generally outperforms inventory keeping at multiple locations (see e.g. Cherikh (2000)). A concentration of wholesalers should, however, be executed slowly, because of cultural differences between various wholesalers. Furthermore, to give up ones business will be emotionally difficult for several of the (previously independently-owned) wholesalers. The required number of wholesalers and there prospective locations should be determined with care.

Furthermore, centralisation of planning activities could potentially improve efficiency. Orders could be processed centrally and distributed over the remaining wholesalers. This would require wholesalers to stop behaving like competitors. They should cooperate with all subsidiaries to achieve common objectives. For example, incoming orders should be forwarded to the central order processing unit. This cultural change can only be achieved gradually. With centralised planning it will be possible to send glass to a customer from the nearest wholesaler or to send a specific type of glass from the wholesaler that has this type of glass on stock. As a result, trucks can be used more efficiently and costs will decrease.

Secondly, the names of the wholesalers could be changed into the name of the owner. Despite the high marketing costs, this process will increase transparency of the supply chain for the customers. Thirdly, a lack of communication is found between wholesalers and the factory and among wholesalers.

Clearly, the suggested changes would impact this supply chain dramatically. Therefore other mechanisms, or only partial implementation of these suggestions, could be desirable. For example, improvement of communications by itself could potentially achieve some of the same benefits as concentration and centralisation. A common computer system could be used to introduce "virtual warehousing". Each wholesaler could then check the stock levels at all locations, transfer orders to other wholesalers, combine shipments and so on. Landers et al. (2000) report possibilities for substantial savings on resources when using this concept.

3.2 Supply Chain of Mobile Phones

Customer demand for mobile phones is still increasing world-wide. As a result, mobile phones tend to be scarce goods at the moment. It often occurs that retailers in The Netherlands cannot offer some of the popular models of major brands like Ericsson and Nokia. Generally, phones are transported from factories to a wholesaler. This wholesaler distributes the phones over various retailers. Most retailers keep only small inventories due to a lack of space.

One of the largest retailers in the Netherlands preferred to be no longer dependent on the wholesaler. Furthermore, one of the major suppliers of mobile phones realised that valuable information, like sales figures, was lost if there is no direct contact between supplier and retailer. Therefore, this supplier decided some time ago to start a pilot project to eliminate the wholesaler from the supply chain and to have direct contact with the retailer. The retailer now orders directly from the supplier and the supplier delivers the order directly to the retailer. The supplier guarantees timely delivery to the retailer, such that the retailer can always offer mobile phones of this brand. Other brands are still delivered to the retailer through a wholesaler.

The supplier takes the responsibility for all logistics activities. However, no knowledge and experience of distribution was available at this supplier. Therefore, the transport was outsourced to a third party logistics provider. Although logistical costs were not analysed, the common opinion among supplier and wholesaler was that a third party logistics provider is cheaper than using a wholesaler. The profit coming mainly from economies of scale in transport and from the fact that products will not be stored in a warehouse.

However, a disadvantage of using a logistics service provider appeared to be the loss of control over timely delivery and the increased probability that phones may be lost or stolen during transport. At this moment, the co-operation is evaluated and it is examined if the co-operation should be continued in this way. One option considered is to return to the concept of a wholesaler, but to retain the direct order process from retailer to supplier. This could then be combined with dedicated shipments from the supplier, which the wholesaler is allowed only to ship to this retailer and not to other retailers. The wholesaler would serve as a cross-docking centre, i.e. ship the products to the retailer on the same day they are received. This could circumvent some of the problems encountered with the logistics service provider, while keeping the same information flows and the same speed of delivery.

Another valid option would be to revise the contract with the logistic service provider. As described in Lim (2000), it is important to draft the contract such that the logistics service provider is stimulated to tell the truth about its capabilities beforehand, and to try to achieve maximum performance afterwards. This could, for example, be achieved by a penalty schema or a gain-sharing scheme.

3.3 Supply Chain of Design Furniture

Furniture is basically sold in two types op stores. Firstly, the stores selling regular furniture, which is produced in large quantities and is available for shipment to the customer on fairly short notice. Secondly, there are showrooms for design furniture. In the latter type of stores, customers can choose from furniture developed by designers and produced by firms that have bought the design. For many types of design furniture, the customer can personalise the product by specifying, for example, the desired height of a couch or the colour of the leather used.

Design furniture is a clear example of a make-to-order environment; a piece of furniture is only produced after the order of a customer is received. Consequently, large lead times exist within this supply chain. The store determines a preferred delivery date based on the wishes of the customer and previous experiences with the factory. Thereafter, the order is sent to the factory by mail or fax. The factory confirms the order and indicates the expected delivery date, which is based on its production schedule and the availability of raw materials. Note that this construction can imply that the expected delivery date is quoted by the factory only after a delivery date has been agreed upon with the customer.

Inventories of raw materials are held at a warehouse. Small replenishment orders are sent at a regular basis to suppliers of raw materials to keep inventory levels low, because furniture is fashion sensitive with the risk of raw materials becoming obsolete. As a consequence, the lead time of the supplier of raw materials influences the delivery date of the furniture as well.

To ensure that the piece of furniture can be delivered to the customer at the agreed delivery date, some safety time is incorporated when negotiating delivery dates with a customer. As a consequence, the piece of furniture is often available at an earlier time than estimated. In such cases, the customer is contacted by the store and it is asked if the piece of furniture can be delivered earlier. If this is not possible, the piece of furniture has to be stored until the agreed date of delivery. Associated costs, for example, for renting storage space at a warehouse, have to be paid for by the factory or store. On the other hand, late delivery of the piece of furniture also occurs, due to delays in the production process of the factory or the supplier of raw materials.

It is clear that the cycle times are high in this supply chain. Furthermore, due to a lack of good communication between the members of the supply chain, delivery dates are difficult to forecast, which results in higher costs. Solving these problems is a difficult task since the supply chain has a complex structure and the members of the supply chain are reluctant to change. The common opinion is that there is no need for changes since there are few complaints.

Delivery times could possibly be improved by using Internet or EDI. If orders are processed over the Internet by using standard forms instead of fax or mail several advantages can be obtained. For example, up to date information on delivery times becomes available instantaneously, communication between members of the supply chain will improve and fewer errors will be made. This can reduce cycle times because the ordering goes faster and because there is no more need to overestimate delivery times. Secondly, by introducing a tracking and tracing system

waiting times can be reduced. The factory and the store can obtain direct information on the status of the order.

Internet could also be used to give the customer the opportunity to send the order directly to the factory. It might however not be expected that the store disappears in this supply chain. Due to the fact that customers probably want to see the furniture in reality before buying, the store can retain the function of showroom. If customers send their orders via the Internet straight to the factory, good agreements have to be made on commission fees for the store. Otherwise, the store cannot survive in this supply chain. See, for example, Griffith and Palmer (1999) for a discussion of the role of intermediates in a supply chain where a manufacturer introduces Internet sales.

It has to be noted that the introduction of an IT application in this supply chain may encounter several difficulties. In Lammings et al. (2000) a distinction is made between supply chains for innovative and unique products on one hand and supply chains for functional products on the other hand. Clearly design furniture falls into the category of innovative and unique products. The paper notes that there are possibilities for sharing large amounts of non-strategic information in a supply chain for innovative and unique products. However, they note that sharing of sensitive information and knowledge is much more problematic than in the supply chains for functional products. Therefore, it seems to be advisable to focus on IT-enabled ordering and tracking and tracing at first.

3.4 Supply Chain of a Supplier of Technical Parts

In this supply chain technical parts are delivered from a wholesaler to a broad range of firms like chemical firms, oil industry and shipbuilders. Products are transported from the factory to the wholesaler. Customers order their products at the wholesaler. The policy of the wholesaler is to have all products directly available. As a result, large inventories are held at the warehouses. According to the wholesaler, the inventories can be divided in fast movers (25%), medium movers (15%) and slow movers (60%). From an economic perspective the amount of slow movers could be reduced. However, according to the wholesaler his position in the market depends on his short response times. Still, many options exist to reduce stock levels while simultaneously improving customer service levels, for example, by improving (or introducing) customer demand forecasts. See e.g. Perry and Sohal (2000). Otherwise, costs and service should be weighed against each other.

By studying this specific chain, it could be concluded that the supply chain is, except for the high inventories, functioning rather efficiently. The customers are satisfied with the delivery times and the quality of the products. During the last ten years, the wholesaler has constantly been busy with improving its business. This business process improvement is used to make processes efficient, effective and adjustable (see also Harrington, 1991). Within the supply chain small improvements are made continuously by using, for example, feedback of employees, suppliers and customers and by applying new techniques, like a new computer system.

3.5 Reverse Flows in a Supply Chain of a Food Retailer

From the distribution centre trucks transport food to various supermarkets. In the distribution centre orders from the stores are collected by order pickers and put into carts. For each store the order is collected in at least 20 and at most 32 carts per truck. Each day a number of trucks arrive at a supermarket according to a truck plan. Within the supermarket the carts are emptied by putting the product in the shelves. To ensure that the order picking process at the distribution centre functions efficiently, sufficient carts should be available. Therefore, empty carts should be returned as soon as possible from the stores to the distribution centre. Some time ago, it was decided within this supply chain that empty carts from one truck should be returned with the next truck that arrives at the supermarket.

For this specific food retailer with 32 subsidiaries, 4050 carts are available. That is, 126 carts are available per subsidiary. On average 25 carts are transported per truck. Consequently, five trucks per store can be filled with carts. However, during a day carts for three trucks are used at most at the same time: carts used by order pickers at the distribution centre, carts in the truck on the way to the supermarket or distribution centre and carts at the supermarket. From this data it could be concluded that sufficient carts are available to ensure an efficient order picking process. However, in practice it appears that the number of carts available at the distribution centre is insufficient.

As a result, return flows of carts should be coordinated more effectively or more carts should be available. By observing three supermarkets and sending questionnaires to all supermarkets, several reasons were found for the fact that the amount of carts available for the order pickers is not sufficient. Firstly, for most subsidiaries it is not clear that there exists a problem at the distribution centre and that they are required to send the carts back with the next truck. As a result, fast emptying of carts has no priority.

As described, trucks arrive at the supermarket according to a truck plan. At Thursday, Friday and Saturday many trucks arrive at the supermarket because of the fact that customers do most of their shopping on Friday and Saturday. The times of delivery are determined by the distribution centre without deliberation with the supermarkets. As a consequence, trucks arrive at moments during the day that there is not enough staff available to empty carts. Furthermore, trucks arrive in such short intervals that there is no time to empty all carts before the next truck arrive. As a result, not all carts are sent back to the distribution centre within a reasonable amount of time. Due to a lack of communication on, for example, truck plans between the distribution centre and supermarkets less carts than required are returned, which results in a problem for the order picking process at the distribution centre.

A third problem is that carts are used at supermarkets for temporary storage of products. On average four carts are used for this per supermarket. It is being attempted to reduce this to two carts per supermarket.

Apparently, sufficient carts seem to be available within this supply chain. However, due to a lack of knowledge of rules for returning carts, lack of communication, lack of staff at supermarkets and the fact that carts are used for other pur-

poses, carts are not returned in time. The solutions for these problems seem to be evident. Communication of rules for returning carts should increase. Motivating supermarkets to send their carts back by introducing a competition among all supermarkets could work positively. Furthermore, supermarkets and the distribution centre should communicate on the truck plan. Wishes of both parties should be balanced. Furthermore, the supermarket is responsible for hiring staff for emptying carts at the right moment of the day, which is based on the truck plan. Finally, permanent racks and shelves should replace carts which are used for this purpose at supermarkets.

If there are still insufficient carts available at the distribution centre after introducing these measures, new carts could be bought. The costs of new carts do not weigh against the costs of a distribution centre that is not functioning.

3.6 Supply Chain of Decorative Coatings

Within this supply chain a chemical firm within The Netherlands produces decorative coatings for large retailers, professional painters and other customers. From the distribution centre the tins with paint are delivered directly to large retailers where customers can buy the paint. From the distribution centre tins of paint are also transhipped to wholesalers, that sell the paint to special stores and professional painters. As described, the paint is sold to customers in tins. Therefore, the chemical firm needs to have an upstream supplier of tins.

Within this specific chain for decorative coatings a backwards integration between the chemical firm and a supplier of tins has occurred. This integration was part of an Efficient Consumer Response project between both firms. The goals of this project are: lower and more reliable inventories within the supply chain, more reliability in the delivery of tins of paint, higher service levels and lower costs within the supply chain. First, we will discuss the specific aspects of the backwards integration in more detail. Thereafter, we will observe if the mentioned goals are realised.

In the project, the main aspect of integration concerns integration of inventories. Before the start of the project, both firms had an inventory of tins. Every month a prediction of the number of tins required was sent to a supplier of tins. In case the supplier had less inventory than ordered then the supplier produced new tins. Because of lack of trust and communication the level of the value of the inventory within the supply chain was equal to approximately 450000 Euro.

To obtain a more efficient supply chain both firms decided to use the principle of Vendor Managed Inventory to obtain an integration of inventories. The supplier is in the general case of Vendor Managed Inventory responsible for the management of the inventories of his product at each location within the supply chain. The supplier of tins manages the inventory at the warehouse of the chemical firm. This process is supported by EDI.

The chemical firm determines, by using historical data, forecasts and production cycles of the supplier, the amount of safety stock and the value of maximum usage. The inventory level at the chemical firm should be at least equal to the safety stock. The chemical firm is responsible for the registration of the inventory.

At the end of each day the level of the stock is sent electronically to the supplier. If this level is lower than the value of the safety stock plus maximum usage, a replenishment order is required. The supplier takes care of this order. As a result, small orders are frequently delivered to the chemical firm.

By observing this new way of managing the inventory within the supply chain, it can first be concluded that the inventories within this part of the supply chain are reduced with 50 percent. Secondly, forecasts of future needs are made more often and they are more precise. Furthermore, the new electronical order procedure is faster and less error prone. The availability of information has increased and as a result efficiency of the production process has increased. Furthermore, unexpected changes in demand can be met by using the safety stock.

It can be concluded that the goals of the project are met. The inventory level is reduced and the availability of tins has increased. Secondly, the reliability level of the production process and the service level have increased. The number of times that there are not enough tins available, are reduced and the possibility to react on fluctuations in the demand has increased because of the fact that processes at both firms are tuned to each other. Finally, the operational costs are structurally decreased. Less skilled staff can be used at both firms and the amount of money invested in inventory has decreased. Furthermore, the supplier of tins has established that it is only supplier of tins for the chemical firm.

One interesting difference remains with standard theory. In this case, the chemical company decides on the safety stock levels, whereas it would have been logically to transfer full responsibility to the supplier. The supplier can then find the best balance between inventory investment, the required frequency of delivery and the expected losses from shortages, provided that a penalty scheme is set up for shortages. Only if the supplier has the freedom to determine all parameters, an efficient process can be determined. See e.g. Cetinkaya and Lee (2000) or Chaouch (2001) for models to optimise this problem. Nevertheless, the project is considered to be a success and will be extended to other subsidiaries of the chemical company.

4 Conclusions

The world of shopping is changing constantly. Nowadays, products can be bought at stores or via the Internet. Customers order more frequently and in smaller amounts. Furthermore, they want their products to be delivered as fast as possible. As a result, firms face a situation where the number of orders increase while the size of orders decreases. To remain competitive and to be able to react promptly on the wishes of customers, harmonisation of all processes in the supply chain is required.

In this paper a brief description is given of some trends that are effecting supply chains. The importance of cycle time reduction, e-commerce, concentration and centralisation of management functions, third party logistics and global logistics is explained. Furthermore, it is indicated how logistics functions can be changed to deal with the trends mentioned.

Supply chains within The Netherlands have been examined to observe in which way various trends are influencing them. In general, it can be concluded that considerable attention has been paid by companies to cycle time reductions. Furthermore, it has been tried to reduce costs and inventory and to implement e-commerce. Each supply chain has chosen its own way to deal with its changing environment.

It can be concluded that progress has been made in harmonising management within supply chains and in obtaining efficient and effective supply chains with satisfied customers. However, the finish has not been reached and probably will not be reached in the near future. Supply chains have to be improved continuously to remain competitive and to be able to fulfil the wishes of the customers.

Acknowledgment

The authors thank their students Eric Arnoldussen, Erik Berkelaar, Ninja Borsje, Sven Broekhuizen, Marieke Copini, Rachel Gans, Cyrille Geeratz, Leon Hoek, Mark Hogewoning, Robert de Jong, Lydia Jonker, Rolf Jan Keijer, Bob Knoester, Noor Mulder, Marc Nobel, Rehana Reyers, Afkenel Schipstra, Angelique Tensen, Michiel Tibboel, Tom Verplancke, Jacqueline van der Voet and Reinoud Willemsen.

References

Carter, P.L., Carter, J.R., Monczka, R.M., Slaight, T.H., Swan, A.J. (2000): The future of purchasing and supply: A ten-year forecast. The Journal of Supply Chain Management, Vol. 36(1), 14-26.

Cetinkaya, S., and Lee, C.Y. (2000): Stock replenishment and shipment scheduling for vendor-managed inventory. Management Science, Vol. 46(2), 217-232.

Chaouch, B.A. (2001): Stock levels and delivery rates in vendor-managed inventory programs. Production and Operations Management, Vol. 10(1), 31-44.

Cherikh, M. (2000): On the effect of centralisation on expected profits in a multi-location Newsboy problem. Journal of the Operational Research Society, Vol. 51, 755-761.

Dornier, P.P., Ernst, R., Fender, M., Kouvelis, P. (1998): Global Operations and Logistics, Text and Cases. John Wiley and Sons, New York.

Fleischmann, M. (2001): Quantitative Models for Reverse Logistics. Lecture Notes in Economics and Mathematical Systems, Volume 501, Springer Verlag, Berlin.

Fraser, J., Fraser, N., McDonald, F. (2000): The strategic challenge of electronic commerce. Supply Chain Management: An International Journal, Vol. 5(1), 7-14.

Griffith, D.A., and Palmer, J.W. (1999): Leveraging the web for corporate success. Business Horizons, Vol.42(1), 3-10.

Hagdorn-van der Meijden, L., Van Nunen, J.A.E.E. (2000): Informatie- en communicatietechnologie en de rol van het distributiecentrum in vraaggestuurde netwerken. Praktijkboek Magazijnen/Distributiecentra, Kluwer, Deventer, 1.4-01 – 1.4-14.

Handfield, R.B., Nichols Jr., E.L. (1999): Introduction to Supply Chain Management. Prentice Hall, New Jersey.

Harrington, H.J. (1991): Business Process Improvement, the breakthrough strategy for total quality, productivity and competitiveness, McGraw-Hill Inc. San Jose.

Lammings, R., Johnsen, T., Zheng, J., and Harland, C. (2000): An initial classification of supply networks. International Journal of Operations & Production Management, Vol. 20(6), 675-691.

Landers, T.L., Cole, M.H., Walker, B., Kirk, R.W. (2000): The virtual warehousing concept. Transportation Research Part E, Vol. 36, 115-125.

Lim, W.S. (2000): A lemons market? An incentive scheme to induce truth-telling in third party logistics providers. European Journal of Operational Research 125, 519-525.

Perry, M., and Sohal, A.S. (2000): Quick response practices and technologies in developing supply chains. International Journal of Physical Distribution & Logistics, Vol. 30(7/8), 627-639.

Rogers, D.S., Daugherty, P.J., Ellinger, A.E. (1996): The relationship between information technology and warehousing performance. Logistics and Transportation Review, Vol. 32(4), 409-421.

Van der Baan, C.A.S. (1997): Knelpunten bij uitbesteding: prijsvechten of strategisch partnership? In Logistieke Knelpunten in het Nederlandse Bedrijfsleven, De Koster, M.B.M., Roos, H.B., De Vaan, M.J.M. (editors), 107-120.

Valuing Time in Make-to-stock Manufacturing: Calculating the Limits of Time-based Competition

Joseph D. Blackburn

Owen Graduate School of Management, Vanderbilt University and Visiting Scholar, INSEAD

Abstract. Although response time is a critical dimension of competition in supply chains, the range of strategies employed in practice varies widely with respect to speed. Some organizations are actively compressing time in the supply chain while others within the same industry are making strategic decisions that embody slower, not faster, response time. The objective of this study is to outline a methodology for valuing time in make-to-stock manufacturing to help clarify the costs and benefits of changing the speed of response. Two important properties of the marginal value of response time are established: (1) the marginal value of time increases as response time is decreased; (2) for equal response times, the marginal value of time is greater at non-optimal inventory levels than at the optimum. We establish limits to time-based competition that demonstrate condition under which slower response may be preferred. To illustrate the methodology, we examine the strategic decisions about response time faced by a manufacturer of truck components.

1 Introduction

Response time continues to play a pivotal role in operations strategy. Through "time-based competition" (Blackburn (1991), Stalk & Hout (1990)) many firms have achieved dominant industry positions by exploiting an ability to respond faster to customers. Much of the literature on this subject chronicles corporate success stories sharing a common theme that "faster is better." Intuitively appealing as that message is, a closer examination of the strategies employed in make-to-stock (MTS) manufacturing reveals an apparent contradiction: a substantial number of U.S. corporations are successfully pursuing strategies that embody slower, not faster, response time.

This time paradox is clearly evident in the textile and apparel industry. The Quick Response (QR) movement was launched with considerable fanfare in the 1980s by the US apparel industry as a strategic response to offshore competition. QR proponents claimed that domestic US manufacturers could offset their higher production costs by responding faster and accurately to the taste changes of fashion-conscious consumers, reducing the huge costs of markdowns and lost sales. QR was called a *"a triumph of information technology, speed, and flexibility over low labor rates."* (Abernathy et al., 1999).

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

The critical question is whether using information technology, speed, and flexibility to enact a quick response strategy can actually triumph over low labor rates. An assessment by *The Economist* in April 2000 stated that *"The epitaph for America's textile and garment industry was written decades ago. Clothes making is labor-intensive... ... each week comes news of another factory closure, the jobs sent to Latin America or Asia where they belong... as go garments, go textiles because fabric making 'follows the needle'.* Since that assessment, more US domestic manufacturers have closed their facilities and moved manufacturing to Asia or Latin America, lengthening the supply chain and slowing response time. Yet other US apparel manufacturers have chosen to stay and use speed to compete with offshore manufacturers.

Apparel is not the only industry to exhibit conflicting trends in response time. The supply line for some components in the automotive supply chain, such as seats, has increasingly become more compressed in time and space. On the other hand, production of some components such as wiring harnesses has been moved to facilities further away from the auto assemblers, usually to lower wage maquiladoras in Mexico. Outsourcing is being widely adopted; some organizations have offloaded the entire production process to contract manufacturers, dilating the response time within the supply chain.

These examples of opposing response time strategies beg further analysis because they appear to challenge the assertion that fast response dominates slow. Upon closer examination these examples force refinement of that claim by highlighting a critical difference in the role that time plays between MTS manufacturing and make-to-order (MTO). In MTO and service systems, faster response time is a key dimension of competition because it is *transparent* to the customer; all other things being equal, customers will choose the MTO firm with faster response; speed wins at the margin. But in MTS, response time is largely invisible to customers. If the desired item is in stock, delivery to the customer can be instantaneous. Shorter replenishment times are unequivocally better than longer only if factors of production such as labor cost are equivalent. For a firm considering a shift of production to a location that extends the supply chain and increases the replenishment time, lower labor costs and other factors of production can clearly impose limits on time-based competition within the supply chain. To understand these limits, tools for evaluating the value of changing response time (faster or slower) in MTS are needed.

The primary purpose of this paper is to develop a methodology for valuing time in MTS manufacturing and its supply chain, where the range of strategies employed in practice varies widely with respect to speed. To make an informed strategic choice, managers must not only know the *direction* of change in costs, they must also know the *magnitude* of how costs change with respect to time.

We use simple, well-known models to evaluate the monetary effects of both increasing and decreasing response time in US manufacturing and apply the models to the case of a U.S. MTS manufacturer, Springfield Manufacturing, who faced strategic response time decisions. The concerns that prompted the study, however, grew out of numerous experiences by the author with organizations that were considering a transformation to lean manufacturing with faster response time and lower cost. These MTS manufacturers had common characteristics: Management

was convinced that the firm's survival against offshore competition depended upon rapid convergence to quick response manufacturing and that massive process improvement initiatives were required to do so. Management was uncertain, however, of the actual value of benefits accruing from faster replenishment times, particularly the value of time; they were only certain that change was needed. Typically, these MTS firms used order quantity/reorder point (Q,r) inventory management policies, but did not maintain their inventories at optimal levels.

2 Literature Review

The effects of changing the leadtime have been studied by a number of authors. In MTO and service settings, response time is a dimension of competition that affects customers' choices, and a number of authors have examined the effect on competitive equilibrium of changes in an organization's response time. A recent study by Lederer and Li (1997) summarizes and extends this stream of research. Hariharan and Zipkin (1995) examine the problem of extending the "demand leadtime"—the time between the due date and when the order is placed— and show that this has an effect equivalent to reducing the replenishment time in a MTS situation. In MTS manufacturing, several studies make the replenishment leadtime a design variable, find the leadtime that minimizes a cost function and describe the properties of the optimal policy. Lovejoy and Whang (1995) investigate response time design by separating the decision into the selection of an order processing leadtime and a production cycle. Dada and Mehta (1996) develop an optimization model for response time as a function of the investment in information technology and then extend the model to include price as a decision variable. In a style goods setting, Matsuo (1990) develops estimates for the marginal value of time by calculating the expected change in cost by producing a product family one time unit earlier. Fisher and Raman (1996) examine the effects of reduced uncertainty through improved forecasting and delayed replenishment decisions for fashion goods with long leadtimes. Hill and Khosla (1995) consider the optimal leadtime in manufacturing, given a specific, demand-dependent cost function for leadtime reduction. They develop an analytical expression for the marginal value of leadtime reduction that is similar to the one we develop, but they do not analyze its properties or extend it. Our study builds and expands upon one of the Hill and Khosla models.

This study differs from the existing literature in several ways that are important in practical situations. With the exception of the paper by Fisher and Raman (1996), the previous studies are theoretical studies of optimal leadtime reduction under the assumption that reduced leadtimes can only be obtained by incurring a higher, well-defined cost. We do not attempt to capture the cost of changing the replenishment time or its affects on quality and productivity. Obtaining a leadtime reduction is a complex organizational process involving, in various measures, information technology, changes in planning procedures, process improvement on the shop floor, and modified logistics. There are costs of implementation, of

course, but the experience of many lean manufacturing initiatives shows that, when changes in quality and productivity are factored in, the reduction in response time may actually result in lower costs.

The models we develop for valuing time specifically do *not* make the restrictive assumption that inventories are being managed at optimal levels. As such, this is a sharp departure from most research on inventory or supply chain management (an exception is Corbett and Alfonso (2000)). We relax the restriction for greater generality because, in our experience, inventory managers are more likely to "satisfice" than optimize. Those rare organizations sufficiently sophisticated to attain optimal inventory levels are likely to have also carried out process improvement efforts to remove most of the time delays from their processes. The models developed in this study are applicable to both groups including those struggling with long replenishment leadtimes, imprecisely-managed inventories, and a vague understanding of the value of time.

3 Problem Description

We study the problem in the context of strategic decisions about response time faced by Springfield Manufacturing, a first-tier supplier of components to a truck manufacturer. The generic problem confronting the firm in our study is presented in Figure 1. At the existing replenishment leadtime L_0 and current level of demand for one of their products, the cost of production and distribution is $TC(L_0)$. Management is considering two radically different strategic options: one would *reduce* the replenishment time to L_1; a second would *increase* it to L_2.

Option 1: The component supplier can take several actions to shrink the leadtime. Investments in information technology can reduce planning cycles and order processing; lean manufacturing programs can compress throughput time on the factory floor; improvements in logistics can speed the transfer of product from manufacturing to distribution centers; moving manufacturing closer to customers can also reduce shipping time. As a MTS operation, the underlying demand is not significantly affected by replenishment time. We assume that, as replenishment times are reduced, finished goods inventories are adjusted to keep customer service objectives invariant. If the firm reduces the replenishment time to L_1, then the cost of production and distribution will be reduced to $TC(L_1)$ (not considering the cost to the organization of making the leadtime reductions).

Option 2: The supplier can also increase the replenishment leadtime to L_2. Of course, this does not mean deliberately slowing down the process, but moving production to a more distant location to obtain lower factors of production (specifically, lower labor cost). As replenishment times are increased, the organization seeks to manage their inventories to keep customer service constant. $TC(L_2)$, the cost of production and distribution with the slower replenishment time, may be higher or lower than $TC(L_0)$ or $TC(L_1)$, depending on the situation.

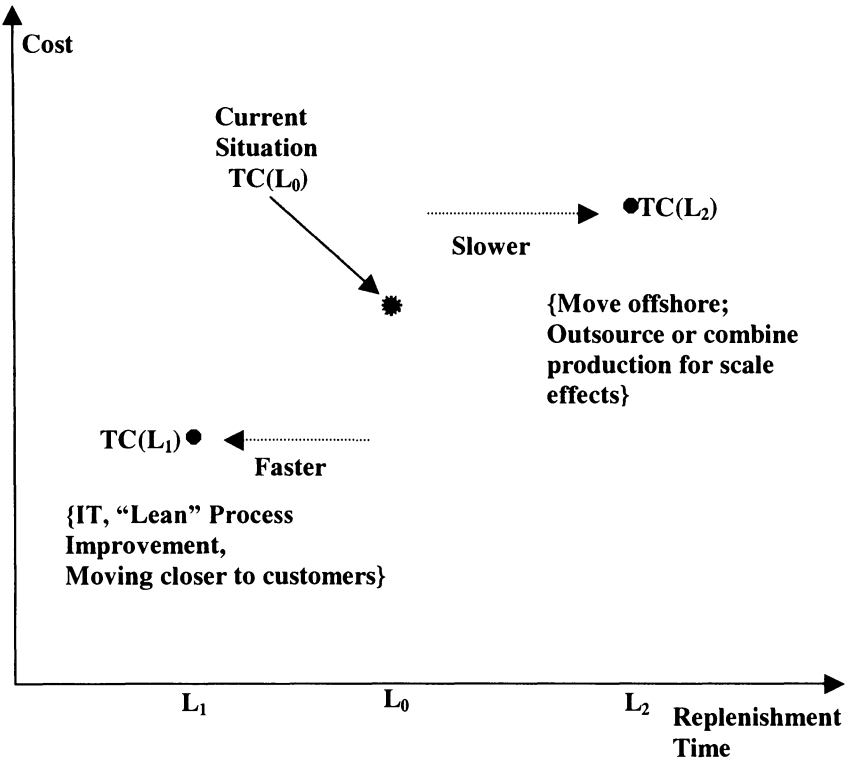


Fig. 1. Strategic alternatives

The objective of this study is to develop a methodology to define the function that expresses the relationship between replenishment time and the cost of production and distribution—that is, to quantify the value of response time.

4 Model and Analysis

We assume that a MTS manufacturer plans production and manages finished goods inventory for a number of products. The inventory for each product is managed by a conventional reorder quantity/reorder point (Q,r) policy.—that is, when the inventory position (on hand+ on order) falls below a level r, a production order of size Q is released to manufacturing. Demand per time period is characterized by a distribution with mean μ and variance σ^2. Demands in different time periods are assumed to be independent and identically distributed, so demand over the known replenishment leadtime, L, has mean $L\mu$ and variance $L\sigma^2$. The firm manages inventory to specific customer service goals driven by forecasts of demand; there-

fore, the reorder point r is set equal to $L\mu + k\sqrt{L}\sigma$ (mean + k standard deviations of demand) for each product. Unfilled orders are backlogged.

Lotsizes Q are computed based on periodic production cycles and, except for possible adjustment of lotsizes, manufacturing capacity is not an issue. For greater generality we specifically do not assume that Q and r (or k) are set at their optimal values, only that the firm maintains a consistent inventory policy (constant k) as the replenishment time is changed; hereafter the policy will be denoted by (Q,k). Quality and productivity are assumed to be invariant with changes in replenishment time L. In the model that follows, the replenishment leadtime is taken as a key decision variable.

Additional notation for the model is as follows:

S = Order cost;

D = annual demand;

c = Unit cost of item;

h = Inventory holding cost/unit;

s = Shortage cost/unit;

P = production/shipment "pipeline" in time units (P = L);

f_p= fraction of finished goods value assigned to product in "pipeline" ($0 < f_p = 1$);

$F(\mu, \sigma^2, L, k)$ = expected quantity short during replenishment leadtime;

G(k) = expected quantity short during leadtime for the standard Normal

$$= \int_k^\infty (u-k)\phi(u)du$$

Then the total expected cost of following a (Q,k) inventory policy, on an annual basis, is

$$TC(Q,k) = SD/Q + cD + (\mu Pf_p + Q/2 + k\sqrt{L}\sigma)h + (D/Q)s\,F(\mu, \sigma^2, L, k) \quad (1)$$

The first two terms of the total cost function are the annual order cost and production cost, the next term represents the sum of annual costs of inventory in the production pipeline, cycle stock and safety stock, and the final term is the annual expected shortage cost. In this total cost expression, the "value" of inventory in the pipeline is represented as a fraction of the finished goods holding cost. Shortages are backordered; each unit of backorder is assessed a shortage cost s.

In what follows we will assume that demands per time period are Normally distributed. This assumption simplifies the development of expressions for the value of time and is a good fit in most situations. We have confirmed by calculation using distributions such as the Poisson and Compound Poisson that our general conclusions about the value of time are not dependent on the normality assumption.

Assuming that demand over the leadtime L is Normally distributed with mean $L\mu$ and standard deviation $\sqrt{L}\sigma$, expression (1) simplifies to

$$TC(Q,k) = SD/Q + cD + (\mu_1 f_p L + Q/2 + k\sqrt{L}\sigma)h + (D/Q)sG(k)\sqrt{L}\sigma \quad (2)$$

Expression (2) is a fundamental, well-documented relationship for MTS inventory management (Silver and Peterson, 1985). We summarize the properties that are relevant to the assessment of the value of changes in the response time, L.

Proposition 1. TC(Q,k) is jointly convex in Q and k.

Proof. See Zheng (1992).

Although Zheng and others have developed algorithms for the joint determination of optimal inventory policies, we only assume that the firm has chosen a production quantity Q and service factor, k.

Observation 1. Expression (2) can be written as the sum of a deterministic (EOQ) problem for the order quantity Q and a time-sensitive, "Newsboy" problem[1]:

$$\text{EOQ Problem cost expression} = SD/Q + cD + (Q/2)h \quad (3)$$

$$\text{"Newsboy" cost expression} = (\mu f_p L + k\sqrt{L}\sigma)h + (D/Q)sG(k)\sqrt{L}\sigma \quad (4)$$

Separating (2) into two expressions clarifies the study of the relationship between cost and response time. All the time-related costs are captured in expression (4): this is an analog of the classic "Newsboy" problem in which the cost of carrying inventory, $(\mu f_p L + k\sqrt{L}\sigma)h$ is weighed against the expected annual cost of shortages, $(D/Q)sG(k)\sqrt{L}\sigma$. Expression (4) is used to estimate the marginal value of time, and the limiting value of response time (when response is instantaneous) is given by expression (3). That is, if production and distribution were instantaneous then (4) would vanish, and we would produce to order and eliminate all the costs associated with replenishment times.

Taking the partial derivative of (4) with respect to L yields the marginal value of time,

$$VT(k,L) = \mu f_p h + \frac{\sigma}{2\sqrt{L}}(kh + (D/Q)sG(k)) \quad (5)$$

Expression (5) is the key to valuing time in a MTS supply chain. The expression has been developed by other authors (Hill, 1992) but in the context of optimizing leadtime reduction. In our applications expression (5) is used to evaluate both increases and decreases in response time. Also note that VT(k,L) is a valua-

[1] See Silver and Peterson (1985), pp. 398-405.

tion tool of substantial generality because it is applicable with equal force for production systems with optimal and non-optimal inventory policies.

We observe from (5) that the marginal value of time is a non-linear function of leadtime and service factor k; however, for a given leadtime, the value of time is a linear function of mean demand and the underlying uncertainty level, the order frequency D/Q, and holding and shortage costs. The sensitivity of the value of time to these parameters will be examined more closely in the application of this model to Springfield Manufacturing. In addition, we note the following properties.

Proposition 2.
(i) Total expected cost is concave increasing in L;
(ii) the marginal value of time, VT(k,L) is convex and a *decreasing* function of L;
(iii) Total cost and VT(k,L) are linear, increasing functions of the underlying uncertainty level, σ.

Proof. (i) The total cost expression (1) has the form $\alpha + \beta\sqrt{L} + \gamma L$, a concave increasing function;

(ii) The first and second partial derivatives of VT(k,L) with respect to L are, respectively, negative and positive;

(iii) can be verified by inspection.

Prop. 2 (ii) also has important strategic implications for time-based competitors. As an organization decreases its leadtime, it receives increasing marginal returns. Therefore, the more leadtime is reduced, the greater the cost incentive for further reduction. If firms A and B produce the same product, and firm A has a faster response time, then A accrues greater benefits from an equivalent reduction in leadtime. The slower firm loses ground in cost competition.

The next proposition states an important difference between optimizing and non-optimizing MTS manufacturers.

Proposition 3. For fixed Q and L, let k* be the minimizing value of TC(Q,k). The marginal value of time is a convex function of k that is minimized at k = k*. That is,

$$\left.\frac{\partial TC(k,L)}{\partial(L)}\right|_{k=k^*} \leq \left.\frac{\partial TC(k,L)}{\partial(L)}\right|_{k\neq k^*} \tag{6}$$

$$\frac{\partial^2 VT(k)}{\partial^2(k)} \geq 0 \tag{7}$$

This proposition follows from partial differentiation of (2) and (5).

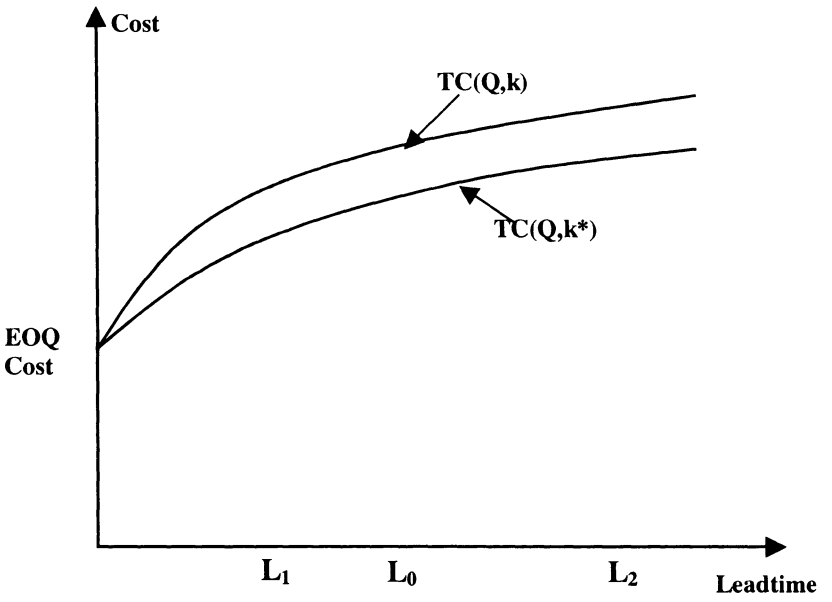


Fig. 2. Time-related costs

Proposition 3 states that, for a given value of response time L, the value of time assumes its *minimum* value when the organization is managing inventories at the optimal reorder point. Although intuition might suggest that organizations managing inventories at optimal levels would expect to gain the greatest cost improvement from response time improvements, the proposition implies that the opposite is true. A firm not managing inventories optimally gains more benefit from time reduction than if they were operating at an optimum level. This should not be viewed, of course, as a benefit to non-optimal behavior, but only that under non-optimal inventory management, the incentive for time compression (and improvement) is greater.

Figure 2 displays how the total cost of production and distribution varies with leadtime for a non-optimal policy, TC (Q,k), and an optimal policy, TC (Q,k*). Leadtime reduction brings increasing, rather than diminishing, returns, but for any given leadtime the marginal value of time is lower for the optimal policy. The two functions converge to the value of the EOQ as uncertainty (and the replenishment time) diminishes to zero. A non-optimizing MTS organization can reduce its costs both by reducing its replenishment leadtime and by improving its inventory management.

5 Modeling The Cost of Increasing Response Time

Although most studies only address the issue of reducing leadtimes in manufacturing, there are attractive strategic alternative that increase replenishment times to access new, cheaper sourcing opportunities. A simple extension of the curves in Figure 2 is insufficient to describe the economic consequences of these longer leadtimes because there is a reduction in the cost of production.

The model that describes this case is the analog of another familiar inventory management model: the "all units" quantity discount model (Silver and Peterson, 1985), except that the total cost is as a function of leadtime, or uncertainty, instead of quantity ordered. This model is displayed in Figure 3: by moving production to a new location (presumably further away) with replenishment time L2, lower costs of production and shipping (c) and holding cost (h) are attained, creating a discount for "all units" ordered. The result is a step function decrease in the cost function at leadtime L2, exactly as in the traditional quantity discount model. As we will show in the application that follows, the actual comparison of total costs for different strategic alternatives is trivial, but yields non-trivial insights for process improvement and location decisions. This model helps establish the limits of time-based competition.

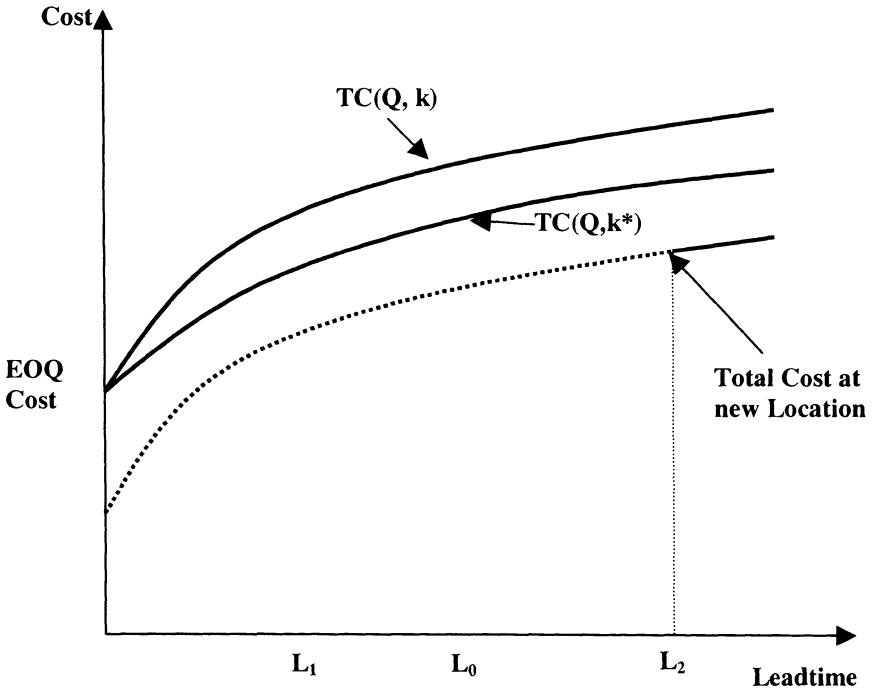


Fig. 3. "All units" time discount model

6 Application: The Value of Response Time at Springfield Manufacturing

Our experience with Springfield Manufacturing illustrates how a MTS manufacturer can use the methodology to evaluate the economic consequences of increases and decreases in response time. Springfield is a typical, small supplier of components in the truck and automotive industry. Their customer—an assembler of large, over-the-road trucks—demands JIT shipments of components in an amount equal to one or two days' demand. Although the assembler provides a general forecast of the overall level of demand for components, the actual orders arrive on short notice for quick delivery. Because Springfield's leadtimes exceed their customer's response time requirements, they are compelled to make-to-stock and fill orders from finished goods inventory. As a small supplier, Springfield has little leverage with which to bargain for advanced information or longer response times; instead, they must continue to reduce costs or risk losing the business to another foreign or domestic supplier.

Springfield supplies up to 100 different specialty components to the truck assembler; these components are primarily exterior trim, constructed of plastic and welded steel tubing (examples are "mud flaps", grille protectors, and aerodynamic trim). For added volume, Springfield also produces a variety of metal and plastic products for other customers, but these products are ancillary to this study.

Like many small manufacturers, Springfield's production facility is organized as a job shop with work areas for cutting, welding, painting, assembly, etc.; most work areas were equipped with sturdy, but aged, machinery. Long setup times at the work centers make large batch production runs necessary, yielding long queues of work and long production throughput times. At the time of this study the total replenishment leadtime for Springfield averaged about 21 days; during the 21 days, we found that there was less than an hour of value-added processing time.

As noted earlier, Springfield's management was considering alternatives to reduce cost and improve profitability. One alternative was faster response: reorganize manufacturing by product families into cells, reduce batch sizes and undertake other process improvements to reduce the total replenishment time from 21 to 7 days. The second alternative was slower response: move production to a Mexican maquiladora to reduce labor cost (by more than 50%), a move that would increase the replenishment leadtime from the current three weeks to about 7 weeks. The threat also existed that another supplier with a lower labor cost offshore facility could bid away Springfield's business.

To evaluate these strategic alternatives, Springfield needed an accurate estimate of the value of changing their replenishment leadtime and its effect on their competitive position. Although in the actual study, we examined the consequences of changing the response time on the entire product mix, only one of their typical products is analyzed here to illustrate the methodology. Data for one of their highest demand components are given below.

Demand averaged about 500 units per week with relatively low levels of variation from week to week; the data were fit reasonably well by normally distributed weekly demand (μ=500, σ=75). As might be expected, there was a small amount of negative correlation between demands in successive weeks, but this was neglected in this study.

The cost of production and distribution was about $15 per unit. Labor content was relatively low, about 6% of that cost. Further study showed that the product could be produced and delivered from a plant in Mexico for about 3.5% of the cost (with the incremental shipping/distribution costs amounting to less than 1%).

Springfield maintained high service levels because frequent shortages endangered their relationship with their customer. Their customer service target was 99%— that is, they carried sufficient safety stock of components in finished goods to satisfy 99% of their customer's demand. Although management was uncomfortable with setting a value on the penalty cost per unit of a shortage, the cost for purposes of the study was estimated to be $10/unit.

Production order quantities were set equal to about four weeks' demand, and components were scheduled to be produced on four-week cycles (2000 units). Since management estimated the cost of setups to be $200, the order quantities closely approximated the value that would be calculated using an EOQ formula. Because of the longer leadtimes, we assumed that a facility located in Mexico would produce in larger batches equal to about six weeks' demand (or 3000 units).

The following assumptions were made in order to evaluate the effect of leadtime reductions on work-in-process, or pipeline, inventories. Pipeline inventory in the local factory was assumed to be equal to 50% of the $15/unit of finished goods inventory. For the Mexican production alternative, pipeline inventory was assumed to have the same value as finished goods, since a majority of the time in the pipeline would be spent in transit as finished goods.

Under these assumptions, we calculate how Springfield's cost of production and distribution vary with response time and display the results in Figure 4. TC(Q,k) shows how total costs vary at the current facility as leadtimes are changed, assuming the customer service policy is maintained at level k. The function TC(Q,k*) shows how total costs vary at the current facility under an optimal inventory management policy. The total cost of supplying product from the Mexican facility is also shown; this function is displayed as a dotted curve because the only relevant point on the curve for this facility is at a seven-week leadtime (the entire curve is shown to illustrate that the marginal value of time is roughly equivalent to the other policies).

There are several important strategic insights concerning the value of time to be made from Figure 4. The horizontal dotted line comparing the cost of shifting production to Mexico (with a seven-week replenishment time) shows that the total costs of production and distribution in Mexico fall between the cost at current operating conditions and the total cost that could be obtained from retaining the 3 week leadtime but moving to an optimal inventory policy. Significantly, the range in total cost across the three policies is quite small: the percentage difference in annual costs between the current policy and an optimal policy, at the same leadtime, is less than 0.5%. Figure 4 shows that, under the current inventory policy, a reduction in leadtime of one week would be required for the total cost to be

equivalent to the total cost of Mexican production with a seven-week leadtime. The marginal change in total cost of reducing the leadtime by one week is about $1200 (which is about 0.25% of the total annual cost).

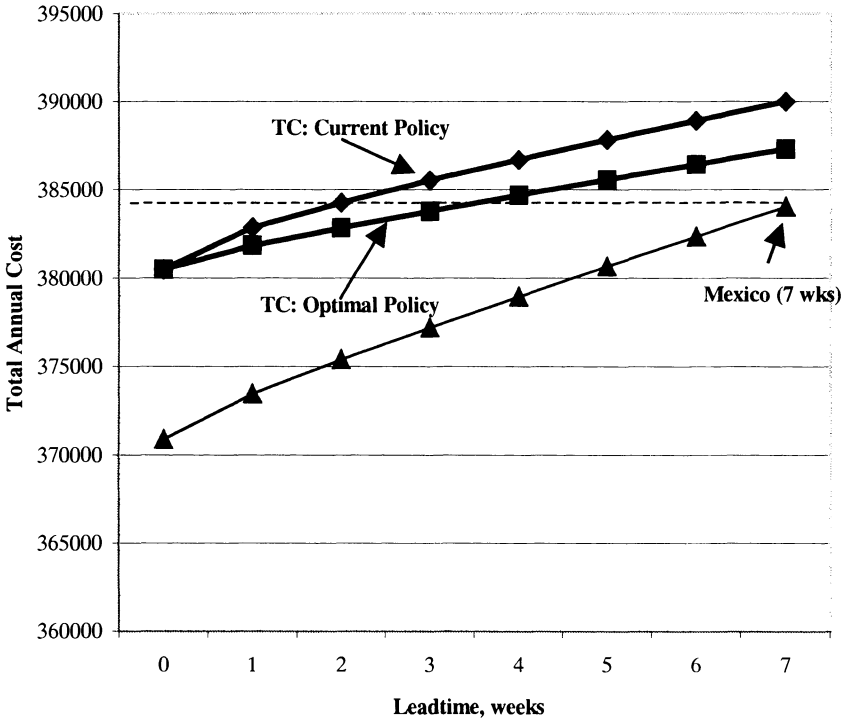


Fig. 4. Springfield manufacturing: Total annual cost vs. leadtime in weeks

Figure 5 provides more insight into the differences among these strategies by showing labor costs and time-related costs as a percentage of total cost for four policies: the current (Q,k) operating policy; a reduction in the leadtime of one week for the current policy; a reduction in leadtime of two weeks; moving production to Mexico. Observe first that moving production to Mexico is essentially cost neutral because the reduction in labor cost is roughly equivalent to the increase in time-related costs; labor savings are traded for time-related costs. It is also important to note that the percentage improvement in time-related costs is surprisingly small: a reduction in the leadtime from three weeks to one week reduces total annual costs only by about 0.6%. This is not to say that a cost reduction of 0.6% should not be vigorously pursued, only that this cost savings may fall short of expectations set by the process improvement literature.

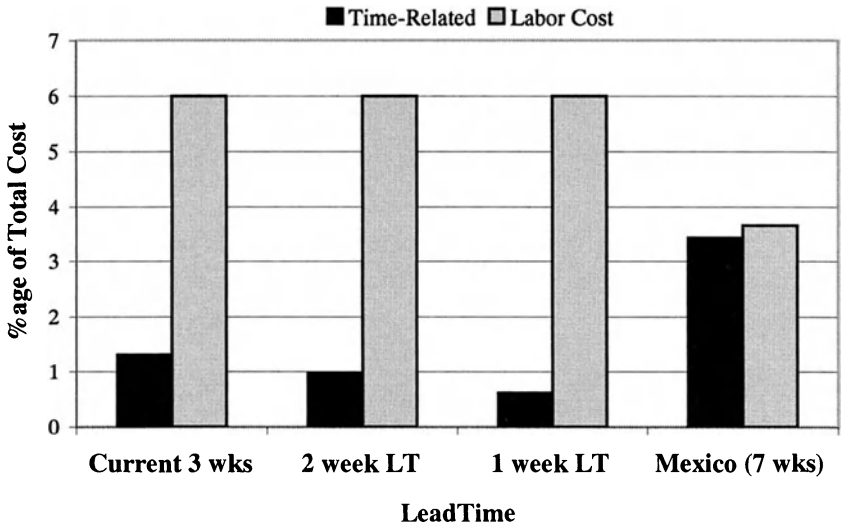


Fig. 5. Time-related costs and labor costs as a percentage of total cost

The results are relatively insensitive to Springfield Manufacturing's specific parameter values. Figure 6 and Table 1 display the results of sensitivity analysis on the underlying uncertainty level, σ, and the shortage cost, s. Figure 6 shows the results if the underlying uncertainty level and shortage cost are doubled. Note that the relative positions of the policies have not changed and that the value of time is relatively insensitive to changes in the parameters. The marginal value of time at the current three-week leadtime increases from \$1213/week to \$1305. One reason for the small change is that as σ changes, the k value to maintain an equivalent service level also changes. Table 1 shows the results for a doubling of both σ and the shortage cost s. Note that the marginal value of time is more sensitive to an increase in the shortage cost, yet the effect on total cost (from 0.34% to 0.44%) is relatively small.

Table 1. Sensitivity of Value of Time

	Marginal value of time (\$/week)	**Marginal value of time as %age of total cost**
3 week lead time (c_v = 0.15; s= 10)	\$1213	0.31%
3 week lead time (c_v = 0.30; s= 10)	\$1305	0.34%
3 week lead time (c_v = 0.15; s= 20)	\$1727	0.44%

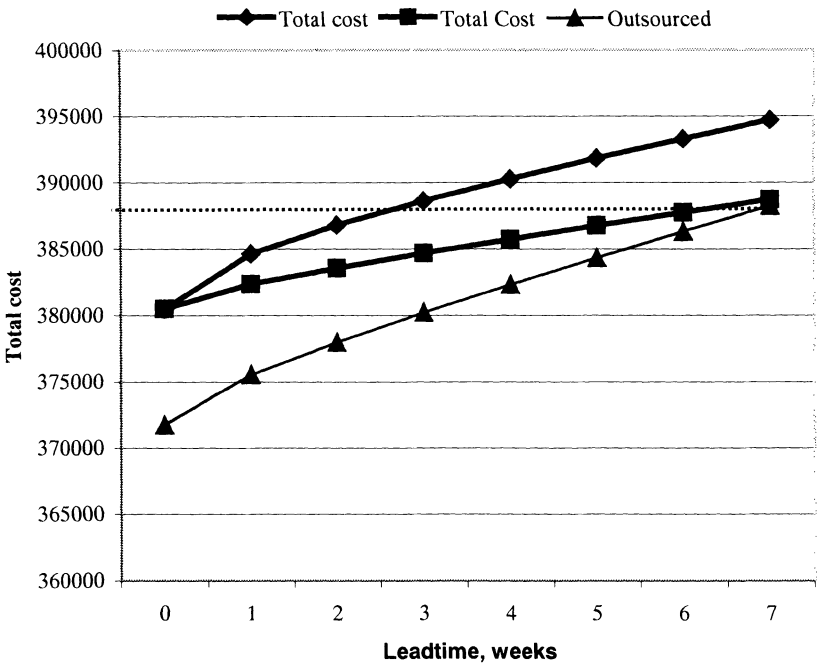


Fig. 6. Springfield manufacturing

Are these observations that a change in leadtime brings a relatively small reduction in time-related costs due to Springfield's cost structure? Our experience with other MTS manufacturing examples suggests that this is not an isolated case; the reduction in total costs due solely to the effects of leadtime decreases on inventory levels tends to be smaller than expected.

What was the outcome of this analysis for Springfield? Armed with a better understanding of the limited benefits of pure time compression, Springfield saw that their current competitive position was perilous and that action was needed. The benefits of reducing leadtime alone were not substantial and would not solidify their competitive position. They recognized that unless there were other compensating benefits—such as quality and productivity improvements—radical restructuring of their manufacturing process would not be worth the effort.

Their initial efforts at process improvement have brought encouraging results. Although leadtimes have only been reduced from 21 to about 15 days, reductions in cost of 1-2% have been achieved across their product line. Most of the benefits have come however from increases in productivity due to faster machine setups and in reductions in scrap and rework; these benefits overshadow the inventory savings that have accrued from the shorter replenishment time.

7 Conclusions

The methodology we have developed gives MTS manufacturers a useful tool for calculating the effects of changes in the replenishment time on production and distribution costs. Unlike most inventory models, our models apply to, but do not assume, optimal behavior, so they have broad applicability. We demonstrate that the non-optimizing firms actually have more to gain from time compression.

We also demonstrate that, absent quantifiable cost improvements in quality and productivity, the benefits of time compression are not as large as generally believed and can easily be overestimated. There are quantifiable limits to time-based competition. Many process improvement activities have failed to produce the benefits expected, and these models help explain why.

Our models also show that the increase in cost from lengthening the supply chain is also less than expected. The results explain why the length of the supply chain in most industries is highly correlated with the percentage of labor cost in the product.

Much more remains to be learned about the value of changing response times. By developing a simple tool that evaluates the costs of inventory and related uncertainty, we have by necessity not modeled the effects on productivity and quality and therefore underestimate the value of speed. To properly evaluate faster response, more comprehensive models are needed to incorporate these other effects.

References

Abernathy, F.H. / Dunlop, J.T. / Hammond, J.H. / Weil, D. (1999): A Stitch in Time: Lean Retailing and the Transformation of Manufacturing—Lessons from the Apparel and Textile Industries. Oxford University Press, New York.

Blackburn, J.D. (1991): Time-Based competition: The Next Battleground in American Manufacturing. BusinessOne Irwin, Homewood, IL.

Dada, M. / Mehta, S.R. (1996): The Impact of Information Technology on Price, Service Level, Response Time and Market Structure. *Krannert Graduate School of Management Working Paper*, Purdue University.

Corbett, C.J. / Alfaro, J.A., (1999): The Value of SKU Rationalization in Practice. *Working Paper, The Anderson School at UCLA*, September, 1999.

Fisher, M. / Raman, A. (1996): Reducing the Cost of Demand Uncertainty through Accurate Response to Early Sales. *Operations Research*, 44, 1: 87-99.

Hariharan, R. / Zipkin, P. (1995): Customer-order Information, Leadtimes, and Inventories. *Management Science*, 41, 10: 1599-1607.

Hill, A.V. / Khosla, I.S. (1992): Models for Optimal Lead Time Reduction. *Production and Operations Management*, 1, 2:185-197.

Lederer, P. J. / Li, L. (1997): Pricing, Production, Scheduling, and Delivery-Time Competition..*Operations Research*, 45, 3: 407-420.

Lovejoy, W. S. / Whang, S. (1995): Response Time Design in Integrated Order Processing/Production Systems. *Operations Research*, 43, 5: 851-861.

Matsuo, H. (1990): A Stochastic Sequencing Problem for Style Goods with Forecast Revisions and Hierarchical Structure. *Management Science*, 36, 3:332-347.

Silver, E. A. / Peterson, R. (1985): Decision Systems for Inventory Management and Production Planning. John Wiley and Sons, New York.

Stalk, G., Jr. / Hout,T.M.(1990): *Competing Against Time*. The Free Press, New York.

Sweatshops to Body Scans (2000): *The Economist*, April 29, 2000, p. 61.

Zheng, Y. (1992): On Properties of Stochastic Inventory Systems. *Management Science*, 38,1: 87-103.

Chapter 2

Reverse Logistics

Internal Pricing in Supply Chains

Kenneth Fjell and Kurt Jørnsten

The Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 Bergen

Abstract. Managing a supply chain concerns environments in which there are multiple decision makers, which may be different firms or different divisions within a single firm. In a supply chain, behaviour that is locally rational can be inefficient from a global perspective. Thus, management attention has to be focused on methods or mechanisms that improve system efficiencies. In this paper we propose a novel negotiated two-part tariff scheme for use by management. It combines earlier results in the literature as well as introduces a reversed use of two-part tariffs in the presence of shortfall in deliveries. We argue that negotiated two-part tariffs can be used for internal pricing as a means to achieve "channel coordination", both under normal delivery and under shortfall, as well as for risk sharing when parties differ in risk aversion.

1 Introduction

A supply chain is two or more parties linked by a flow of goods, information and funds. Since this means that supply chain management concerns environments in which there are multiple decision makers, which may be different firms or different divisions within a single firm, attention has to be focused on methods or mechanisms that improve system efficiencies. The reason for this is that in a supply chain setting, behaviour that is locally rational can be inefficient from a global perspective. One way of achieving improved system efficiency is through contractual arrangements.[1]

In the *first-best* case, total expected supply chain profit is maximised. This can be achieved if all decisions are made by a single decision-maker with access to all available information. This corresponds to a supply chain with centralised control. However, in a supply chain, this situation is highly unlikely to occur since normally none of the parties involved is in a position to control the entire supply chain. Thus we are left with a situation in which the supply chain control is decentralised and where each party involved in the supply chain has private information and individual objectives.

In the case where the total profit encountered through decentralised control is lower than the total profit for the first best solution, the decentralised control mechanism used is regarded as inefficient. There are several objectives that are of

[1] Here, "efficiency" is taken to mean the combined expected profit of the vertical supply chain.

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

importance in the design of the control mechanisms in a supply chain. One objective is risk sharing, where contracts are focused on how the total system profit should be allocated among the supply chain parties. The reason for calling this focus a risk sharing objective is that it provides means for the parties to share the risks arising from various sources of uncertainty.

Other contracts are more focused on reducing the difference between the total profit in a decentralised system and the total profit in the first-best solution. Contracts in which this is the main emphasis are said to focus on system wide performance and the normal terminology for the objective used is channel coordination.

In the excellent survey article by Tsay *et al.* (1999), the means for achieving channel coordination and/or risk sharing are used to classify the literature on supply chain contracts. This classification uses the following categories:

- specification of decision rights
- pricing
- minimum purchase commitments
- quantity flexibility
- buyback or return policies
- allocation rules
- lead time
- quality

In this paper we will focus on internal pricing as a way to achieve supply chain coordination. We will suggest the use of negotiated two-part tariffs to achieve the coordination goal that is in line with the suggestions made by Moorthy (1987) in his comment on the article by Jeuland and Shugan (1983). However, the idea on how the two-part tariffs to be used in the pricing scheme are to be derived is based on an idea by Lantz (2000). In Lantz' negotiation scheme the prices to be used among the parties are decided by negotiations on a two-part tariff. Lantz has shown that this negotiation scheme may converge to a situation in which the seller sets the linear part of the two-part tariff to the marginal cost of production and the fixed fee of the tariff allocates the profit among the participants according to their negotiation power. Hence, the scheme may result in a situation in which the problem of double marginalization is avoided and the profits are distributed in a "fair" way. The reason for using the statement "may converge" is that the results of negotiations are dependent upon the starting conditions, i.e. the internal pricing scheme used in the outset of the negotiations.

2 Pricing as a Mean for Supply Chain Coordination

In the early literature on inventory control, the pricing scheme to be used in order to co-ordinate the chain is not subject to negotiations among the parties involved. However there exists a substantial literature on the use of quantity discounts as a

means to achieve better channel coordination. Monahan (1984), Lee and Rosenblatt (1985, 1986), Parlar and Wang (1995), Lal and Staelin (1984) and Gupta and Kini (1995), just to mention a few. In most of these articles the profit is solely allocated to the seller with the exception of the article by Rosenblatt and Lee (1985), in which both parties benefit from the discount schedule. In the marketing literature, the paper by Jeuland and Shugan (1983) presents a two-member channel and shows how a discount-pricing scheme can be used to achieve channel coordination. However, in a comment to the Jeuland and Shugan paper, Moorthy argues that the channel coordination suggested by the authors can be achieved by the use of a two-part tariff and that this coordination scheme is simpler and hence easier to implement. Also, Moorthy shows that the two-part tariff separates the coordination problem from the profit sharing problem. Since Moorthy's ideas in some sense form the basis for our suggestion, we will further comment on Moorthy's coordination scheme.

In the more recent literature on pricing and supply chain contracts, several extensions of the classical models have been presented. Weng (1995, 1997) extends the Jeuland Shugan model in several directions and shows among other things that a quantity discount scheme for the buyer along with a franchise fee paid to the supplier is sufficient to achieve channel coordination. Hence, Weng suggests a fixed fee element to be used together with a quantity discount scheme. One of the drawbacks with most of the pricing schemes presented in the literature is that most of them focus on a supply chain consisting of two members; a buyer and a supplier.

In the recent paper by Lee and Whang (1999), the authors present a number of properties that are of interest when designing coordination mechanisms in supply chains. These are the cost conservation property, the incentive compatibility and the informational decentralizability property. The cost conservation property states that the accounting system built into a coordination scheme should trace all costs to the individual sites. The incentive compatibility property means that the scheme should eliminate potential misalignment problems. Finally, the informational decentralizability property states that the scheme should be possible to implement using site information only. Using these desired properties, we argue that a negotiated two-part tariff coordination scheme might be the answer. One of the reasons for this is its simplicity, which makes it possible to use in situations where the supply chain consists of more than two-parties and hence most of the coordination schemes suggested in the literature become too complicated or are not applicable.

3 Two-Part Tariff as a Preferred Pricing Scheme

In his comment on the Jeuland and Shugan article, Moorthy states that there are several pricing schemes that can be used to achieve channel coordination. The necessary and sufficient conditions for channel coordination is that the manufacturer's pricing policy makes the retailer's effective marginal cost curve cut the retailer's marginal revenue curve from below at the channels optimal quantity.

Hence, the manufacturer's pricing scheme need not even be a quantity discount scheme. Moorthy shows that by using a two-part tariff in which the manufacturer charges his true marginal cost in the linear part of the tariff, channel coordination is achieved and the fixed fee is used only to determine the profit sharing.

Jeuland and Shugan criticise the two-part tariff suggested by Moorthy on the grounds that it is insufficient in dealing with production uncertainty. Specifically, they argue that if, because of some short-term problem of scheduling production, the manufacturer delivers only a fraction of what the retailer wishes to purchase, then the manufacturer suffers no penalty at all. However, Moorthy states that this criticism is non-valid since it is easy to negotiate a "protection-from-shortfalls-in-production" clause into the contract. Moorthy suggests the following clause "if the manufacturer fails to supply (the ordered quantity) in any month, then the retailer will be entitled to a refund of the fixed payment in an amount sufficient to maintain the retailer's share of the channel's profits ... plus any accrued interest on the excess amount already paid ...." (Moorthy, 1987, p. 378, brackets added).

We suggest that this clause can be replaced by a two-part tariff working in the opposite direction of Moorthy's two-part tariff. Hence, there will be *two* two-part tariffs in place. First a two-part tariff stipulates the selling price. The linear part of this tariff is preferably set equal to the marginal cost of the manufacturer thereby ensuring channel coordination, whereas the fixed fee gives the profit sharing among the parties. This tariff is in use in the normal situation when the ordered quantities are promptly delivered. The second two-part tariff works in the shortfall situation, i.e. to handle production uncertainty. In order to maintain channel coordination, it is important that the linear part reflects the true marginal shortage costs of the retailer. It should thus include implicit costs such as loss of goodwill from unserved customers. This will enable the manufacturer to correctly adjust extraordinary use of resources such as overtime or emergency repairs to remedy the shortfall. The fixed fee in this reversed tariff serves the purpose of redistributing the channel profit among the parties along the lines suggested by Moorthy. However, like the former fixed fee, it will also reflect the negotiation skills/powers of the parties.

Agrawal and Seshadri (1995) have shown that a menu of two-part tariffs can also be used for risk intermediation in a supply chain with risk averse retailers facing demand uncertainty. Unlike the ordinary two-part tariff proposed by Moorthy, the fixed fee is now paid by the (risk-neutral) manufacturer to the retailers as a compensation for demand risk. In addition to assuming more of the demand risk, the manufacturer also assumes responsibility for the ordering decision setting the quantity to maximize expected value of the supply chain. The manufacturer chooses the menu of two-part tariffs to maximize own expected profit. Such a risk sharing mechanism could also be included in our scheme as it consists of two two-part tariffs working in opposite directions.

Next, we discuss a suitable negotiation scheme for promoting efficiency in the supply chain through the implementation of two-part tariffs.

4 Negotiation Process to Reach Channel Coordination

In his thesis, Lantz suggests that negotiated two-part tariff should be used as internal prices. The way the negotiation works is that starting from a given position with a fixed and neutral internal pricing tariff, each of the involved parties can suggest a new two-part tariff. The new suggested tariff is announced to the other party in terms of a fixed fee, F, a linear unit price part, p, and an order quantity, Q. The suggestion of the new tariff is based on the suggesting party's private information. This means that the new tariff is certainly favourable to him in comparison with the former tariff. The responding party can accept or reject the suggested tariff. If the new tariff is accepted it must be favourable also to the other party. The negotiations continue until no further change in the two-part tariff is suggested. A rejection of a suggested tariff means that the parties involved for the moment stick to the old tariff.

Lantz has tested the negotiation scheme by letting students negotiate a two-part tariff in a situation in which the manufacturer has private information on his cost structure and the retailer has private information regarding the demand structure. In his experiments the results are promising and the negotiations lead to a better channel coordination.

We have carried out similar experiments with students in the production and logistics courses at NHH with largely the same type of results. However, when the students start out negotiations based on a linear tariff, the experience is that the chance of progression to a two-part tariff is low. Further, it is our experience that students often perceive the negotiations as a zero sum game, focusing directly on profit sharing, instead of first focusing on channel coordination to maximize combined profits, and subsequently on profit sharing. It is not entirely unlikely that these same problems will surface in real supply chain management negotiations.

5 Conclusions and Future Research

In this paper, we have focused on ways of using internal pricing in the form of two-part tariffs in combination with a particular negotiation process to improve channel coordination. Combining earlier results in the literature with a new reversed use of two-part tariffs to handle shortfall in deliveries, we propose a novel internal pricing scheme for improving channel coordination. Our pricing scheme works as follows. Starting from a neutral position in terms of tariffs used for deliveries and backorders, the parties negotiate as in Lantz' (2000) scheme, but now negotiating *two* different two-part tariffs.

1. Each of the involved parties can suggest a new two-part tariff to be used for deliveries from the manufacturer to the retailer consisting of a fixed fee, F, a unit price, p, and an order quantity, Q. The other party may accept or reject the proposal.

2. Each party may also suggest a new two-part tariff to be used in terms of shortfall. This new two-part tariff consists of a fixed fee, Φ, and a unit shortage price, π, to be used when the ordered products are not delivered promptly.

This system of negotiated two-part tariffs can easily be generalised to a situation where the supply chain consists of more than two-parties. Again, to ensure channel coordination the ordinary linear part should reflect marginal production cost and the reversed linear part should reflect marginal shortage costs of the "retailer". Furthermore, the suggestions for new tariffs are made using private information only and each party can, given the tariffs to be used, base their decisions on the tariffs and their own private information.

In terms of further research, we are currently constructing a negotiation game consisting of a supply chain with three parties where the retailer in the chain faces uncertain demand. This will be used in testing if the participants in the game can reach a situation that co-ordinates the channel based on the two types of two-part tariffs suggested in this paper. Apart from this, we will also analyse how our scheme works in situations were other pricing schemes have been suggested in the literature and try to determine the pros and cons of the suggested pricing scheme.

References

Agrawal V. / Seshadri S. (1995): Risk intermediation in supply chains *IIE Transactions* 32 pp 819-831 2000

Gupta O.K. / Kini R.B. (1995): Is price-quantity discount dead in a just-in-time environment? *International Journal of Operations Management* Vol. 15 No. 9 pp. 261-270

Jeuland A. / Shugan S. (1983): *Managing Channel Profits.* Marketing Science 2 pp. 239-272

Lal R., Staelin R. (1984): An approach for developing an optimal discount pricing policy. *Management Science* Vol 30 No 12 pp. 1524-1539

Lantz Björn (2000): *Internprissättning med effektiva incitament.* Ph.D thesis Handelshögskolan i Göteborg, Bokförlaget BAS

Lee H.L. / Rosenblatt M.J. (1986): A Generalized Quantity Discount Pricing Model to Increase Supplier`s Profits. *Management Science* 30 pp. 1179-1187

Lee H. / Whang S. (1999): Decentralized Multi-Echelon Supply Chains: Incentives and Information. *Management Science* 45/5 pp. 633-640.

Monahan J.P. (1984): A Quantity Discount Pricing Model to Increase Vendor Profits. *Management Science* 30 pp. 720-726.

Moorthy K.S. (1987): Managing Channel Profits: Comment. *Marketing Science.* Vol. 6. No. 4. pp. 375-379

Moses M. / Seshadri S. (2000): Policy Mechanisms for supply chain coordination. *IIE Transactions.* 32 pp 245-262.

Parlar M. / Wang Q. (1995): A game theoretical analysis of the quantity discount problem with perfect and incomplete information about the buyer`s cost structure. *Recherche operationelle* Vol. 29 No. 4 pp 415-439

Rosenblatt M.J. / Lee H.L. (1985): Improving Profitability with Quantity Discounts under Fixed Demand. *IIE Transactions* 17/4 pp. 388-395

Tsay A.A. / Nahmias S. / Agrawal N. (1999): Modeling Supply Chain Contracts: A Review. In S. Tayur et al *Quantitative Models for Supply Chain Mangement.* Kluwer

Weng Z.K. (1995): Channel Coordination and Quantity Discounts. *Management Science*, 41 pp. 1509-1522

Weng Z.K. (1997): Pricing and Ordering Strategies in Manufacturing and Distribution Alliances. *IIE Transaction*s, 29 pp. 681-692

Closed-loop Supply Chains

V. Daniel R. Guide, Jr.[1] and Luk N. Van Wassenhove[2]

[1] Duquesne University, 600 Forbes Ave., Pittsburgh, PA 15241, USA
[2] INSEAD, Boulevard de Constance, 77305 Fontainebleau Cedex, France

1 Introduction

There are numerous examples and cases available of products that are being reused via remanufacturing or recycling, or combinations of reuse activities (Thierry et al.1995, Krikke et al.1999, Guide 2000, Toktay et al 1999, Fleischmann 2000). However, these products and their supply chains are not all the same with respect to a number of critical dimensions, including product acquisition, reverse logistics, inspection, testing and disposition, remanufacturing, and distribution and selling of the remanufactured product. In the following sections we document a number of diverse products that are presently being remanufactured and describe their supply chains. After each case, we summarize and discuss the distinguishing features of the supply chains. Finally, we discuss the management of each of the different supply chain systems, and identify the key research issues.

2 Supply Chains for Refillable Containers

Xerox Copy/Print Cartridge Return Program

Xerox introduced their program for copy/print cartridge returns in 1991 and, at present, it covers 80% of the toner/print cartridge line. In 1998 Xerox expanded the program to include the recycling of waste toner from high-speed copier and commercial production publishing systems. The return rate for cartridges in Europe and North America was greater than 60% for 1998. This equates to over 2.86 million kilograms of material remanufactured or recycled just from cartridges. Xerox reports avoiding almost 23 million kilograms of landfill because of their reuse programs. The cartridges are designed for remanufacturing and recycling of materials not fit for remanufacture.

Customers return the cartridges by placing the spent cartridge into the packaging used for a full cartridge and attaching a prepaid postage label provided by Xerox. The returned cartridges are cleaned, inspected, and then parts are reused or materials recycled. The full cartridges are then distributed through normal distribution channels to customers. The final cartridge product containing re-

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

manufactured parts or recycled materials is indistinguishable from cartridges containing exclusively virgin materials. Figure 1 shows the supply chain, in simplified form, for a cartridge. Xerox is presently testing the use of 'ecoboxes' to allow bulk returns from high volume users in Europe. Xerox will arrange for regular pick-ups of the boxes by its own carrier network. A bulk returns process allows each high volume user to batch cartridges and may lower the returns costs absorbed by Xerox (the information in this section was obtained from Xerox 1999).

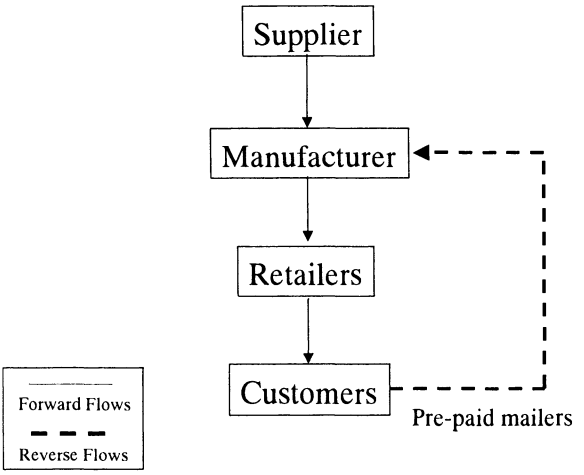


Fig. 1. A Closed-loop Supply Chain for Cartridge Reuse

Kodak Single-Use Cameras

Kodak started their program to reuse their single-use camera line in 1990. The first stage of the program was to re-design the cameras so that parts could be reused and film reloaded. The entire line of single use cameras may be remanufactured or recycled and the amount of materials per camera that are reusable range from 77-80%. The second stage involved forging agreements with photofinishers to return the cameras to Kodak after consumers had turned them in for processing. Kodak now enjoys a return rate greater than 70% in the United States and almost 60% worldwide. Since 1990, Kodak has reused over 310 million cameras, and has active programs in over 20 countries.

The process flow for the reuse of cameras, after the sale of the camera, starts with the consumer returning the camera to a photofinisher to develop the film. The photofinisher then batches the cameras into specially designed shipping containers and sends them to one of three collections centers. Kodak has entered into agreements with other manufacturers (Fuji, Konica, and others) of single-use cameras that allow for the use of common collection centers. At the collection center, the cameras are sorted according to manufacturer and then by camera model. Af-

ter the sorting operations the cameras are shipped to a subcontractor facility where the cameras are cleaned of packaging materials, disassembled, and cleaned. Some parts are routinely reused, some are removed (batteries) and the frame and flash circuit board are carefully tested. These sub-assemblies are then shipped to one of three Kodak facilities that manufacture single use cameras. At the Kodak facility, the cameras are loaded with film and a fresh battery (flash models only), and finally new outer packaging. The final product is now distributed to retailers for resale. The final product containing remanufactured parts and recycled materials is indistinguishable to consumers from single use cameras containing no reused parts. Figure 2 shows the supply chain network for reusable cameras (the information in this section was developed from Kodak 1999).

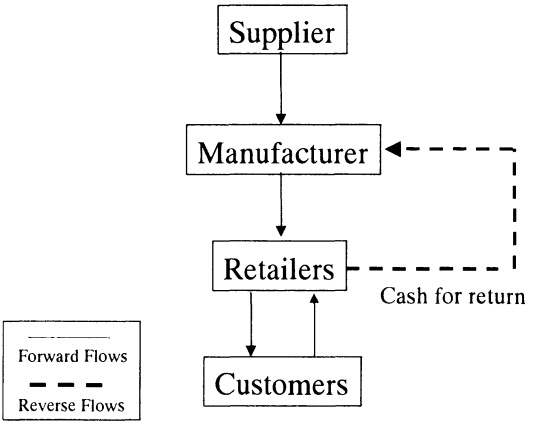


Fig. 2. A Closed-loop Supply Chain for Single-Use Cameras

3 Characteristics of Refillable Container Closed-loop Supply Chains

These products, toner/printer cartridges and single-use cameras, are both containers required to sell the contents. Consumers do not distinguish between a new and a remanufactured/reused product since the purpose of buying the container is to gain access to the contents, which are toner ink and film, respectively. Additionally, the customer also has no way of determining, aside from labeling, whether the container has been remanufactured or that it contains reused materials. The returned containers in both cases are blended with new materials so there is no distinguishing between new and reused products. This is possible because the technology is extremely stable for toner/print cartridges and single-use cameras. The markets for the remanufactured products are exactly the same as for the new products since the two are indistinguishable. The characteristics of the supply chain

for container reuse are listed in Table 1. The volumes are very high, in the millions for both products, and the annual returns quantities are stable or have a known growth rate. In both cases the OEM controls the reverse logistics network and the reuse alternatives.

Table 1: Characteristics of Closed-loop Supply Chains for Refillable Container

Characteristic
Commodity goods
Containers for consumables
High volume
Low variance
Non-distinguishable products
Simple products
OEM controlled
Short lead times

4 Supply Chains for Industrial Remanufacturing - Copymagic

Industrial remanufacturing is mechanical item remanufacturing and accounts for the majority of remanufacturing operations in the United States (Guide 2000). Xerox reports that over 90% of their copier equipment is remanufacturable and that using remanufactured parts has enabled Xerox to reduce landfill materials by over 300 million kilograms since 1998 (Xerox 1999). Xerox has redesigned its photocopier lines using modular design principles to allow for part and component reuse.

Thierry et. al (1995) describes in detail the product recovery processes used by a multinational photocopier manufacturer (nicknamed 'CopyMagic') and we present an overview of the reuse processes and the product characteristics. CopyMagic brings new products to markets and takes back used products from customers. The majority of the return flows are from expired lease agreements, although some used products may be purchased to ensure sufficient quantities of used products. After products are returned there are several alternatives for product reuse: repair, cannibalization, recycling, and remanufacturing. A product may be used for any or all of the above reuse options. For example, a product may be disassembled and some parts cannibalized for use as spares, other parts may be repaired and reused, some other parts fully remanufactured, and parts worn beyond higher-order reuse recycled. The most appropriate level of reuse for products, components, and parts should be driven by economics, e.g. the most revenue-effective reuse option.

CopyMagic does all remanufacturing operations in-house and uses a common production line for both new and remanufactured products. The use of existing production facilities greatly reduced start-up costs for remanufacturing, but has the disadvantage of complex production planning and control. Remanufactured prod-

ucts may be offered as technical upgrades, or with original technology, and this has enabled CopyMagic to exploit additional market segments. Product design relies heavily on modular design concepts to allow for inter-changeable parts and components between models and product lines, in addition to allowing technical upgrades. Products designed for remanufacturing are also easier to repair and service, allowing for better customer relations. The firm has been able to reduce the number of suppliers because of common design platforms and components. A disadvantage to remanufacturing is that suppliers with better quality parts may have lower sales volumes since their parts may be reused more frequently.

The marketing of remanufactured products is complex and CopyMagic has had to tightly control the quality of remanufactured products, and convince customers that the quality of remanufactured products is the same as a new product. However, the company has an enhanced image due the 'green image' of reused products.

The additional complexities of product reuse has caused CopyMagic to design and implement new information systems to forecast and track product returns, analyze the returns for yields and the condition of products, modules, and parts, and to track the performance of remanufactured units. The supply chain (illustrated in Figure 3) required for these activities is more complex to plan, monitor and control. In Figure 3 we show the remanufacturer as being physically separate from the manufacturer. In the case of third party remanufacturing this is standard practice. However, Xerox and Océ both use a common facility for new and remanufactured products. Where a common facility is used, new and remanufactured parts are mixed to produce end items. In this case the distinction between manufacturing and remanufacturing is not physical, but simply to illustrate the different flows of components and parts.

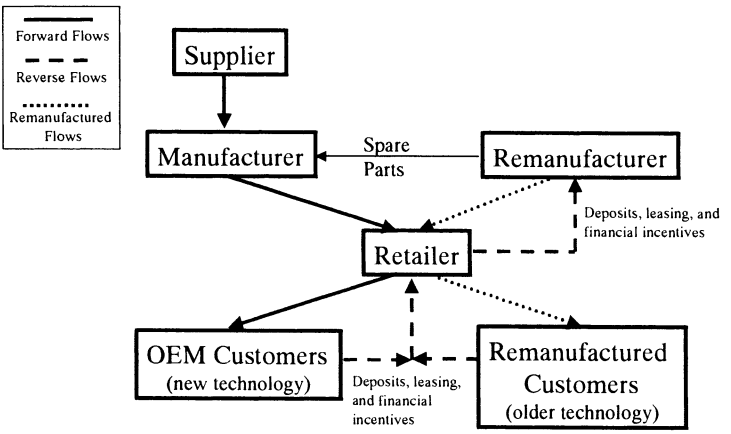


Fig. 3. A Closed-loop Supply Chain for Photocopiers

5 Characteristics of Closed-loop Supply Chains for Industrial Remanufacturing

Xerox and CopyMagic are in the minority in industrial remanufacturing since less than 5% of remanufacturing in the United States is done by OEMs (Guide 2000). Additionally, both Xerox and CopyMagic rely heavily on leasing to enable the firms to forecast the timing and quantities of product returns. Product returns from leasing, for non-OEM remanufacturers, represents less than 5% of total returns and this makes the tasks of forecasting product returns timing and quantities much more difficult. This forecasting problem manifests itself as product imbalances since matching return rates with demand rates is complex and difficult. OEMs have another distinct advantage in the area of product design, since most products must be designed for reuse, i.e., a modular design and clearly labeled materials. The volume of returns in the case of value-added remanufacturing is significantly less than for container reuse.

Marketing is more complex for the remanufactured products since customers may require significant education and assurances to convince them to purchase remanufactured products. Market cannibalization is also a significant concern since little is known about how remanufacturing sales affect new products sales.

Table 2: Characteristics of Closed-loop Supply Chains for Industrial Remanufacturing

Characteristic
High variances
Stable production technology
Limited volumes
Modular design
Imbalances in supply and demand
Cannibalization

6 Closed-loop Supply Chains for the Reuse of Consumer Electronics

Our last case discusses the reuse of consumer electronics equipment and details the supply chain and remanufacturing operations.

Cellular Telephone Reuse at ReCellular, Inc.

ReCellular, Inc., was founded in 1991 in Ann Arbor, MI by Charles Newman to trade new, used, and remanufactured cellular handsets. The business grew from a venture that provided cellular telephones for leasing and alternative sources for handsets were required to reduce costs (hence the discovery of the used handset market). ReCellular is a trading operation that refurbishes cellular phones when

necessary to add value for existing orders, and buys and sells wireless handsets of all technologies. At present, ReCellular estimates it has remanufactured over 1.3 million cellular phones. One of the goals of the company is to be the "first in the second" in the wireless exchange plan. The company offers remanufactured (refurbished) products as a high quality, cost effective alternative to new cellular handsets. Customer services include: grading and sorting, remanufacturing, repackaging, logistics, trading, and product sourcing (all services are specific to cellular handsets and accessories). ReCellular operates globally with a presence in South America, the Far East, Western Europe, Africa, the Middle East, and North America. The company has plans to expand operations to provide better coverage throughout the world.

The cellular communications industry is a highly dynamic market where the demand for telephones changes daily. Demand may be influenced by the introduction of new technology (e.g., digital and analog), price changes in cellular airtime, promotional campaigns, the opening of new markets, churn (customers leaving present airtime providers), and the number of new cellular telephone manufactured. Additionally, there is no worldwide standard technology (e.g., Europe uses GSM, but the United States does not support this wireless technology) that necessitates dealing in a number of often-disparate technologies and standards. These global differences makes regional activities difficult since there may be no local market for certain types/models of phones, requiring a firm to manage global sales and procurement. Additionally, cellular airtime providers may limit the number of telephones supported by their system and the dropping of a phone model by a major carrier can greatly affect a local market. These factors make competition for an original equipment manufacturer challenging. However, a company offering used or remanufactured equipment faces numerous factors affecting the supply of used cellular phones. The same factors that complicate demand affect the availability of used handsets. Further, identifying and taking actions to reduce the uncertainties in the reverse flows as early as possible is an important issue in this business. The supply of used handsets is a volatile market, with volumes and prices in a constant state of flux. Supply uncertainty is not a complication faced by traditional OEMs.

In order to fully understand the nature of the market, both forward and reverse flows of materials must be considered. Figure 4 shows the supply chain system for cellular telephone reuse. The forward movement of materials consists of the traditional flows from suppliers to manufacturers, manufacturers to airtime providers (retailers in this case since the sale of a cellular phone is tied to airtime activation) and airtime providers to the customers. The reverse flows are more complex. Remanufacturers of cellular telephones do not collect handsets directly from the end user, but rather rely on airtime providers or a variety of third-party collectors (we discuss the specifics of product acquisition in the next section). Airtime providers and third-party collectors act as consolidators who then broker the units to remanufacturers. ReCellular then sorts and grades the handsets, and sells the handsets as-is or remanufactured to airtime providers and third-party dealers working with airtime providers. Some handsets may be obsolete or damaged beyond higher-order recovery and these products are sold to scrap dealers and recy-

clers (note this flow may come from both remanufacturers and third party collectors). Recyclers recover polymers and other materials in the handset assemblies, and base materials in batteries. Scrap dealers may separate the handsets into materials and resell individual parts for reuse in other applications and offer the other sorted materials to recyclers. Suppliers may then purchase the recycled materials for use in new products.

The acquisition of used telephones is central to the success of a remanufacturing firm. The nature of product acquisitions is driven by what future demands (unknown) will be for phones. The lead times for delivery after used phones have been purchased are often lengthy and subject to a large amount of variability. This has caused remanufacturers to have stocks on-hand to compete for sales. ReCellular obtains used phones in bulk from a variety of sources, including cellular airtime providers and third-party collectors.

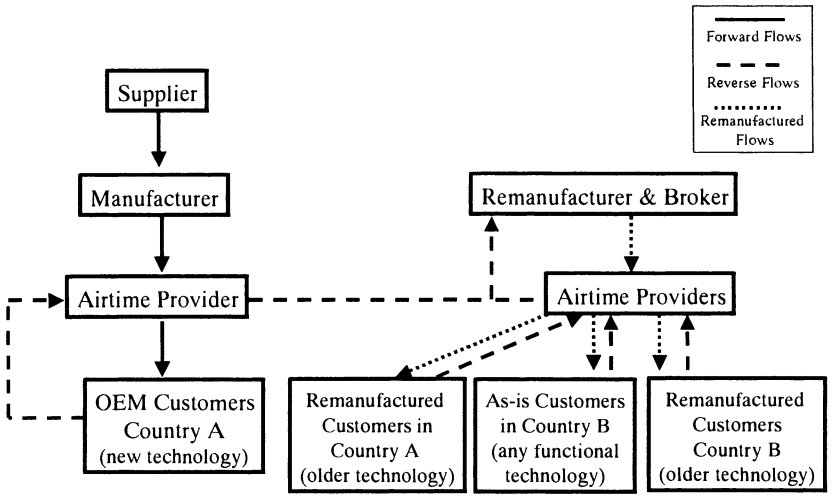


Fig. 4. A Closed-loop Supply Chain for Cellular Telephones

Third-party collectors are often charitable.foundations (e.g., the Wireless Foundation: http://www.wirelessfoundation.org) that act as consolidators by collecting used handsets and accessories from individuals. Cellular airtime providers also act as consolidators by collecting used phones from customers who have returned the phones at the end of service agreements, or customers upgrading to newer technology. Both these and other sources worldwide may offer a variety of handsets and accessories in varying condition for a wide range of prices and quantities. Due to the low cost (approximately $0.50 per phone using air transport) of bulk transportation of phones using a worldwide network of suppliers of used phones is practical and cost-efficient. No individual returns are accepted since the channels required for direct returns from the consumer have too high a cost to be effective at this time. Obtaining the best grade of used products for the best price is one of the keys tasks necessary for the success of ReCellular. Deciding on a fair price to offer for the used phones is a difficult and complex task. At present, the acquisi-

tion staff devotes much of their time to identifying reliable and reputable sources of used phones and establishing a working relation with these suppliers. New suppliers usually require a physical visit to ensure the quality of the used phones.

The value of a used handset is highly dependant on future market demand for that particular model either in remanufactured or as-is form. The present demand for a graded as-is used cellular phone or a remanufactured phone is known for that instant in time, but due to the highly dynamic nature of the industry, these prices are not stable. The market forces discussed earlier may cause the value of a particular model of phone to drop or rise with little warning. An additional factor is that the selling price for remanufactured phones tends to drop over time, making the used phones a perishable product.

This nature of the product reuse market necessitates a fast, responsive supply chain that identifies sources of used phones for a fair price, and future buyers of these phones in either graded as-is or remanufactured condition. Additionally, the system must procure the phones in a timely manner, sort and grade the telephones, have the capability to rapidly remanufacture the phones to order, and provide a fast accurate transportation method to ensure timely delivery of the phones. Re-Cellular is developing an e-commerce site strictly for business to business in order to facilitate matching suppliers and buyers of equipment. The present e-commerce site shows the current stocks (model, price, grade, and quantity) available and what models are needed (but not the prices offered). The site is being upgraded to allow real-time transactions by sales and procurement agents. Future considerations also include using the e-commerce site to facilitate on-line auctions for used and remanufactured products.

7 Characteristics of Closed-loop Supply Chains for Consumer Electronics Reuse

The volume of cellular telephones in use worldwide is enormous, with over 55 million cellular subscribers in the United States alone (U.S. Central Intelligence Agency 2000). One of the first requirements for a remanufacturer in this environment is global coverage. Since the rate of technology diffusion is different for each country in the world, phones, which may be technically obsolete in Norway, may be current technology in Ecuador. This imbalance in the diffusion of technology makes having global operations and intelligence crucial for a profitable operation. Tightly tied to the concept of global markets is the problem of acquisition, or obtaining the best-used product for the best price. The prices for cellular telephones are highly dynamic and are based on future expected prices for remanufactured handsets. The problem is further complicated by the market for graded, as-is cellular telephones where a selling is price is known, as opposed to an expected price in the future for a remanufactured item.

Cellular telephones are perishable items because of the high clockspeed in new product development. Electronics industries have the highest clockspeed, an average of 18 months, and this makes responsive systems crucial to making a profit.

Remanufacturers cannot afford to remanufacture to stock in this environment since the value of the items drops daily.

Table 3: Characteristics of Supply Chains for Consumer Electronics Reuse

Characteristic
Dynamic spot markets for supply and demand
High volumes
Perishable good
Cascade reuse opportunities (worldwide market)
High information requirements
High variability

8 Management of Closed-loop Supply Chains

In the simplest terms all closed-loop chains do have a common set of activities. The recovery process consists of several highly inter-related sub-processes: product acquisition, reverse logistics, inspection and disposition (consisting of test, sort and grade), reconditioning (which may include remanufacturing) and distribution and selling of the recovered products. However, the previous cases illustrate that while there are common processes, not all closed-loop supply chains are alike. Each different supply chain system has different characteristics and management concerns. Table 4 highlights the differences in the three forms of closed-loop supply chains we discussed in this chapter. In the following subsections we discuss each of the sub-processes for the different types of closed-loop supply chains.

Table 4: Key Distinctions Between Closed-loop Supply Chains

	Product Acquisition	**Reverse Logistics**	**Test Sort Grade**	**Recondition**	**Distribution And Selling**
Refillable Containers ⇒ Toner cartridges	Easy	Easy	Easy	Easy	Easy
Industrial Remanufacturing ⇒ Copiers	Intermediate	Hard	Hard	Hard	Hard
Consumer Electronics Reuse ⇒ Cellular Phones	Hard	Intermediate	Easy	Easy	Intermediate

Refillable Container Closed-loop Supply Chains

Product acquisition refers to product acquisition management, and is actually a number of related processes (Guide and Van Wassenhove 2000). First, product acquisition management determines whether reuse is a value-creation activity for a

specific firm. Second, if reuse activities are profitable, to maximize revenue the appropriate method for managing product returns should be selected. Third, operational issues, such as facility design, product planning and control policies, inventory policies, are dependant on the method selected to manage product returns. Fourth, product acquisition management activities help identify and develop new markets for reused product, and to balance the return rates with market demands. We are concerned here primarily with the second and fourth product acquisition processes, selection of the appropriate method for managing product returns, and balancing return rates with market demands. In the case of single-use cameras, the OEM controls the returns process by using cash incentives to motivate the photofinisher to return the empty cameras to Kodak's reuse facility. Xerox also directly controls the returns process for print/toner cartridges. Xerox supplies each customer with a pre-paid mailer and appeals to customers to send back the spent cartridge. Both strategies are extremely successful at ensuring constant volumes of containers, and are examples of market-based returns strategies. Market-based strategies are active strategies to encourage product returns, in contrast to waste stream systems where returns are passively accepted from the waste stream (Guide and Van Wassenhove 2000). Both firms enjoy stable returns flows with predictable volumes each period.

Balancing return and demand rates in a refillable container closed-loop supply chain is a relatively easy process, in part because the customer cannot differentiate between reused and new products. We do not mean that the processes involved in balancing return and demand rates are simplistic (see Tokay, et al. 1999 for a complete discussion of this process), but rather that the process is simple in comparison with the other types of closed-loop supply chains. The technology contained these products is stable and there are very limited secondary markets available for the used containers. There are secondary markets for the refilling of printer cartridges and single use cameras, these are small localized operations that are considered more of a nuisance since the remanufacturing may be sub-standard and damage the firm's reputation. The returned products are mixed with new (replacement) materials as needed and then repackaged and sent back out through traditional distribution channels. There is no need to segment demand by product type (remanufactured vs. new) or for a manufacturer to consider market cannibalization.

Reverse logistics activities are the processes required to move the products from the end user to the facility where reuse activities will take place. In both examples of refillable containers, these sets of activities are simple. The photofinisher acts as a consolidator for Kodak and eliminates the need for Kodak to deal with individual customers. Xerox has minimized their contact with end-users by using a pre-paid mailer, which the customer may then use to arrange for pick-up and transportation. Fleischmann (2000) provides a complete discussion of reverse logistics networks and their characterization.

The disassembly, test and inspection processes and the remanufacturing processes for refillable containers are simple. The product itself is simple and the costs are low, products that may be questionable may be recycled with little or no concern about replacement materials.

Finally, the distribution and selling processes are simple since traditional distribution networks are used and the customer base is the same as for new products.

Industrial Remanufacturing Supply Chains

We summarize the key factors for success in Table 5. Of the closed-loop supply chains, the industrial remanufacturing sector is most likely the best documented, but the most difficult to plan, manage, and control.

Product acquisition activities for industrial remanufacturing may be based on leasing, in which case the product will be either returned at the expiration of the lease, or the lease renewed with the same item. Product acquisition in the case of leasing may be viewed as relatively simple, however, only a small number (5%) of industrial remanufacturers report using leasing (Guide 2000). Remanufacturers report using a number of other techniques (deposits, rebates, and cash refunds) with varying degrees of success and this indicates that the problem of obtaining sufficient quantities is more difficult than for refillable containers.

Table 5: Key Factors Industrial Remanufacturing

Keys to Success – Industrial Remanufacturing Closed-loop Supply Chains
• Ability to forecast and control the timing, quantity, and quality of product returns (IS)
• Design for reuse (recycle, repair, and remanufacture) (engineering)
• Customer education (remanufactured as good as new)
• New relationships with suppliers (fewer parts & components, and design)
• Complex production planning and control problems

Balancing return and demand rates for industrial remanufacturing is more difficult since there are distinct and separate markets for new and remanufactured goods. Customers may require the newest technology or may simply perceive remanufactured products as inferior. These separate markets may affect the distribution and selling of remanufactured products. Market cannibalization may be a concern for manufacturers providing remanufactured products and, as a result, the retail markets may be disparate. The reverse logistics process is most often very difficult since the remanufacturer must arrange for pick-ups from many geographically diverse facilities. Many used industrial items are also regarded as hazardous waste and must be treated as such during transportation. The process of testing, sorting, grading and inspection is time consuming and complex, with the possibility of a single product containing tens of thousands of parts and components. The screening process must be rigorous since products of poor quality may be expensive to remanufacture, or dangerous to reuse. The remanufacturing/reconditioning processes are most often complex and difficult to plan, manage and control (Guide 2000).

Consumer electronics reuse

These types of closed-loop supply chains may hold the greatest promise due to the volume of products available for reuse, but at the same time these types of supply chains represent some of greatest challenges. We list the key factors for success in Table 6.

Table 6: Keys to Consumer Electronics

Keys to Success – Consumer Electronics Closed-loop Supply Chains
• Ability to forecast and control the timing, quantity, and quality of product returns in a global market (IS)
• Fast response (perishable items)
• e-Commerce to identify buyers and sellers
• Identify and exploit cascade reuse
• Identify and exploit technology diffusion differences

Product acquisition is very hard for this form of closed-loop supply chains. The products are used globally, but the rate of technical diffusion is different in various geographic areas. This requires that a successful operation will have worldwide collection and distribution markets and these markets will not be the in the same geographic areas. Supply and demand rates and prices are extremely volatile. The products are also perishable items since the value of a remanufactured item may drop daily because of the rapid rate of technological progress and the rate of technology diffusion. There are also multiple options for reuse since products may be sold in graded as-is condition or remanufactured. Each option has a different selling price which is quite dynamic.

However, there are several of the major processes that may be characterized as easy to intermediate. The nature of the products, very few mechanical parts, makes them simple to test, sort and grade, and remanufacture/recondition. The reverse logistics processes are somewhat hard to coordinate since there are so many national borders with customs regulations to manage. However, the handsets are small and light and may be shipped in bulk with commercial air carriers for an inexpensive price (approximately $0.50 as of the summer of 2000). The distribution and selling processes involve a number of different nations, and requires knowledge of the cellular technology in use and the airtime providers. The selling process is, for reasons discussed previously, tightly intertwined with the acquisition process.

9 Conclusions and Research Issues

There are a number of unique structures for closed-loop supply chains. These different structures all require a set of common activities: product acquisition, reverse logistics, test, sort and grade, remanufacturing /reconditioning, and distribution and selling. The successful management of the various activities does not always

involve the same actions from supply chain to supply chain. It is crucial for managers and researchers to understand the differences and the implications of these differences. One of the basic research needs is a continuing refinement of the documentation, categorization, and understanding of the different forms of closed-loop supply chains. Other pressing research needs are those activities shown in Table 5 to be 'hard'. Finally, we hope that holistic models will be developed that reflect the complex nature of the interactions.

References

Fleischmann, Moritz (2000), *Quantitative Models for Reverse Logistics*, Rotterdam, The Netherlands: Erasmus University Rotterdam Ph.D. Thesis.

Guide, V. Daniel R., Jr. (2000), 'Production planning and control for remanufacturing: Industry practice and research needs', Journal of Operations Management, 18: 467–483.

Guide, V. Daniel R., Jr. and Luk N. Van Wassenhove (2000), Managing Product Returns for Remanufacturing, Fontainebleau, France: INSEAD Working Paper 2000/35/TM/CIMSO.

Kodak (1999), 1999 Corporate Environmental Report, Rochester NY: The Kodak Corporation.

Krikke, Harold, A. van Harten and P. Schuur (1999), 'Business case Océ: Reverse logistics network re-design for copiers', OR Spektrum, 21: 381-409.

Thierry, Martijn, Marc Salomon, Jo Van Nunen and Luk N. Van Wassenhove (1995), 'Strategic issues in product recovery management', California Management Review, 37: 114-135.

Toktay, Beril, Lawrence Wein and Stefanos Zenios (2000), 'Inventory management of remanufacturable products', forthcoming in Management Science.

U.S. Central Intelligence Agency (CIA) (2000), The World Fact Book 2000, Washington, DC: The United States Central Intelligence Agency.

Xerox, (1999), 1999 Environment, Health and Safety Progress Report, Webster, NY: The Xerox Corporation.

Extended Design Principles for Closed Loop Supply Chains: Optimising Economic, Logistic and Environmental Performance

Harold Krikke[1], Costas P. Pappis[2], Giannis T. Tsoulfas[2] and Jacqueline M. Bloemhof-Ruwaard[1]

[1] Erasmus University, RSM/Fac.Bedrijfskunde, P.O.Box 1738, 3000 DR, Rotterdam, The Netherlands
[2] University of Piraeus, Dept. of Industrial Management, 80, Karaoli & Dimitriou Str., 18534 Piraeus, Greece

Abstract. Closed loop supply chains aim at closing goods flows thereby limiting emission and residual waste, but also providing customer service at low cost. In this paper we study design principles for closed loop supply chains, both from a theoretical perspective and a business case. Obvious improvements can be made by applying traditional 'forward logistics' principles. Also new, life cycle driven principles need to be applied. This can be supported by advanced management tools such as LCA and LCC.

Keywords: closed loop supply chains, case-study, reverse logistics

1 Introduction

Over the past few years increasing volumes of return flows, varying from end-of-life returns to marketing or commercial returns, has reinforced interest in the effective management of such flows. More and more, Original Equipment Manufacturers are held responsible by new environmental legislation for the recovery of their own products. In case the OEM is out of the country, importers are held responsible and new parties, mainly profit-oriented, deal also with the recovery and recycling of used products. This results in closed goods flows on a product, component and material level, as shown in Figure 1. A new managerial area called reverse logistics management emerges, which can be described as the process of planning, implementing and controlling the efficient and effective inbound flow and storage of secondary goods and related information opposite to the traditional supply chain, for the purpose of recovering value or proper disposal (Fleischmann, 2000). Typically, this comprehends a set of processes such as collection, inspection/separation, reprocessing (including disassembly), disposal and redistribution (see Fleischmann et. al, 2000).

Closed loop supply chain management goes beyond that. It comprehends all business functions and hence decisions regarding the adaptation of business

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

strategy, marketing, quality management, information systems, logistics and so on in view of closing goods flows, thereby limiting emission and residual waste, but also providing customer service at low cost. Both the forward and reverse chain are considered, since there is a strong interaction between the two.

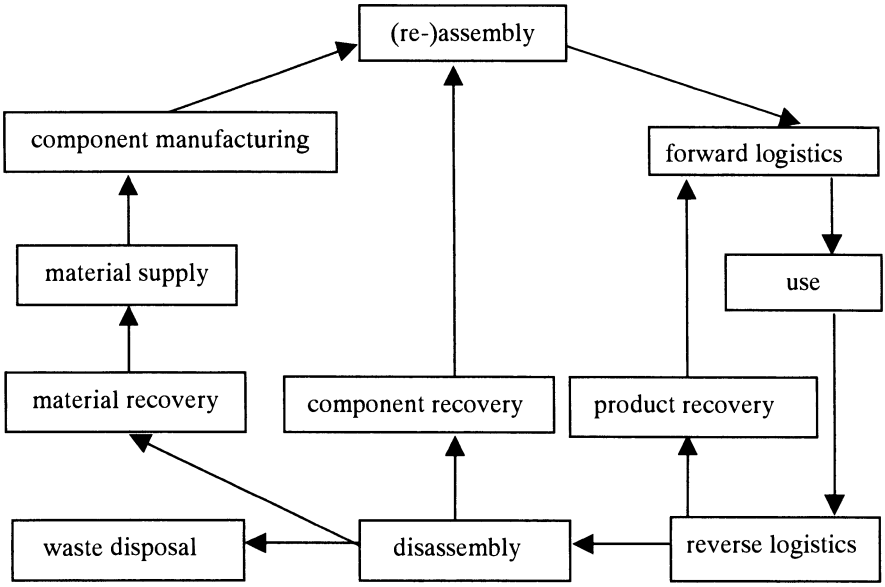


Fig. 1. Forward and reverse supply chain (similar to Ferrer, 1997)

It is essential to analyse in what respect closed loop supply chains fundamentally differ from forward logistics, and how this affects design principles. Closed loop supply chains are different on the following aspects (Fleischmann et. al, 2000) and (Faucheux and Nicolaï, 1998):

- In addition to cost and service there are environmental drivers, complicating the objective function.
- Higher system complexity, in particular in closed loop systems due to increased number of - and interaction between goods flows. Uncertainty on the supply (collection) side of the system regarding volumes, quality, composition and timing.
- Push-pull nature. There is often a mismatch between supply and demand. "Production" (i.e. supply of used products) is not coupled with "demand" (i.e. producer's requirements).
- Numerous "suppliers"/ few "customers". Used products are the raw materials for the reverse chain. Unlike the forward chain, there are a lot more sources of raw materials and they enter the reverse chain at small cost or at no cost at all. However, although obtained for 'free', the value of return flows is low and may be limited to a small fraction of the flow.
- Unexplored market opportunities. Environmental requirements can be the

basis of the creation of new markets or result in the reorganization of existing ones for by-products of the production process. With such reorganization, materials that would otherwise end as wastes would turn into useful products.

The above distinguishing characteristics might justify the development of special approaches. However, little attention has been paid to the question whether design choices in closed loop supply chains differ from those in traditional 'forward' logistics. For example, logistics networks may be more decentralized in closed loop supply chains, but the underlying trade-off between economies of scale and transportation costs might be exactly the same.

The purpose of the paper is to see to what extent return flows influence design principles for (closed loop) supply chains. We ask ourselves the following questions:

- What forward and closed loop logistics design principles are known from the literature?
- Are design principles for closed loop supply chains fundamentally different or do different parameter values simply lead to different solutions?
- To what extent are design principles well understood and applied in business practice?

We develop a theoretical framework in Section 2, which is split in a first part concerning design principles from traditional supply chain literature and a second part concerning design principles from closed loop supply chain studies. In Section 3 we discuss the Honeywell case to illustrate application and understanding of the principles in practice. Clearly, one case is insufficient to get a full picture of what's going on in business practice, however valuable lessons can be learned. In Section 4 we discuss results and draw conclusions.

2 Theoretical Framework

2.1 Design Principles in Traditional Logistics

Several principles, which apply to supply chain (or parts of it) design, are referred to in bibliographies (e.g. Ralph Sims, 1991). A quite comprehensive list of such principles, which appears in (Gattorna, 1997) is described below. Where appropriate, we added the interpretation of these principles for closed loop supply chains. For organisational purposes, we classify the design principles according to the well-known supply chain management areas: organisational aspects (including marketing and strategic issues), information systems, planning and control and network structure. We will use the same classification for the extended principles.

A. Organisational

1. Link logistics to corporate strategy

All aspects of logistics operations must be directly linked to the corporate strategic plan. It appears that many companies do not consider closed loop supply chains management as a strategic issue (Caldwell, 1999). However, companies with successful closed loop supply chains management, such as Xerox, BMW, 3M etc., do consider asset recovery as essential part of their business.

2. Emphasize human resources

Logistics excellence flourishes in an environment that recognizes people as the department's most important resource. Recruiting, education, training and job enrichment are standard practice. Experienced, well-trained managers are critical to the success of business strategies and plans. In returns management, elderly employees, those who assembled the products 5-20 years ago, are often employed in the Asset Recovery department, since they are the only ones with the knowledge necessary to recover the returns.

3. Form strategic alliances

Forming close partnerships with other participants in the product chain or channel can boost logistics operation. Pre- or non competitive R&D is often done in cooperation. For example, Sony, Motorola, Nokia, IKP, Indumetal and Gaiker jointly develop new construction techniques enabling a returned product, once exposed to the correct trigger temperature, to self-disassemble. But also collection and recovery may be done by joint systems.

4. Target optimum service levels

Companies need to calculate their "optimum" service levels and pinpoint the costs associated with sustaining those levels. Clearly, only a few companies see closed loop supply chains management as a service tool. For product lines phased out, return flows may serve as a cheap source of spare parts.

B. Information Systems

5. Use the power of information

Successful logistics implementation takes full advantage of information and information-processing technology, not only for data interchange, but also for decision support. Good return management requires up-to-date information on the installed base, making use of product data management systems, remote sensing and tracking and tracing systems. Also, information can be retrieved from returns, for example by analysing wear out of returned cores.

C. Planning and Control

6. Focus on financial performance

The logistics function should use return on assets, economic value added, cost and operating standards, or similar indicators as measures of performance. Functions as transportation, warehousing and customer service are best managed as cost or profit centres. So far, closed loop supply chains are seen as a cost issue by most companies, however potential revenues from reuse and the avoidance of disposal costs is often neglected.

7. Manage comprehensively

All corporate logistics functions should be unified under a combination of centralized and decentralized management. Grouping all logistics-related functions under a single umbrella facilitates effective decisions. In several studies the authors found that the responsibility for the returns handling was often not clearly assigned, neither in the supply chain nor at a company level.

8. Manage the details

Attention to details can mean real savings. Effective detail management produces consistency of purpose, objectives, image and information to customers. Obviously, this principle is equally applied to both forward and reverse supply chain design and operations. For example, when tracing cause of returns (failure, bad manual, over advertising etc.), it is necessary not to have general figures but to analyse carefully per retailer, distribution channel, type of product and so on.

9. Measure and react to performance

Companies must measure their logistics performance and react to the results in an on-going dynamic fashion. Closed loop supply chains processes should be benchmarked as any other business process.

D. Network structure

10.Leverage logistics volumes

Successful logistics operations consolidate shipment volumes, inventories and the like to gain operating and financial leverage, whether the logistics function is performed in-house or by an outside contractor. Leveraging can be increased by good collection systems and joint ventures. A problem with return flows is that only a small part is valuable and therefore remanufactured/reused whilst the majority is low value and will be scrapped. Hence, reverse substreams follow different recovery routes, which complicates consolidation.

By definition, closed loop supply chains aim at closing goods flows, thereby limiting emission and residual waste, but also providing customer service at low cost. Thus, closed loop supply chains do what traditional supply chains do and in addition contribute to sustainability. Therefore traditional design principles also apply to closed loop supply chains, although in some cases with slight adaptation or different interpretation. In addition, we investigate design principles specifically for closed loop supply chains in 2.2.

2.2 Extended Design Principles

From a closed loop supply chains point of view, the above list of design principles may be extended to include other important rules. Again we remark that both forward and reverse chain are relevant here. From literature study, we are able to propose the following:

A. Organisational

11. Impose sustainability standards on suppliers. Selecting sustainable suppliers requires additional selection criteria. One of the issues to be solved is the supplier paradox: the one supplying reusable parts may loose most business. This needs to be compensated, for example by outsourcing repair to the original supplier, who as a bonus also has most knowledge and dedicated equipment. Also, suppliers may co-design the product to enable modularisation and design for recycling. See also (Corbett and Van Wassenhove, 1993) and (Tsoulfas and Pappis, 2001).

12. Create new markets. The environment can be at the basis of the creation of new markets or of the reorganization of existing ones for certain (material) flows resulting from the production process. With such a technical reorganization, materials that would formerly have ended as wastes are turned into useful by-products (Faucheux and Nicolai, 1998). Furthermore, companies can also offer waste disposal services (Corbett and van Wassenhove, 1993). Companies that manipulate materials and energy should be organized in such a way that they can respond rapidly to changes in management and processes (Tsoulfas and Pappis, 2001). Changing demands for goods and services will also push design changes. The study of alternative plans is necessary in order to achieve eco-optimisation. "Do the same but do it better or try to do something different."(Klassen and Angell, 1998). Pro-activeness, especially to intended legislation, has proven to be effective in many situations.

13. Make use of management tools, such as ISO 9000-14000, life cycle analysis, environmental accounting methods, that may help business to identify and select opportunities for improvement. For example, using less energy is generally good for the environment, but is also self-evidently good for business because it cuts companies' costs, and eventually avoids potential environmental liabilities. It is, therefore, a prerequisite to the long-term sustainability of business. To replace non-renewable and polluting technologies, it is crucial to support the use of solar, wind, water and geothermal energy (among others), as well as reduction in energy consumption (Tsoulfas and Pappis, 2001).

B. Information Systems

14. Make use of accounting systems that account for the full life-cycle costing of a product or service, and the environmental impacts it creates. Based on this, develop and design recoverable products, which should be technically durable, repeatedly usable, harmlessly recoverable after use and environmentally compatible in disposal (Gotzel et. al, 1999).

C. Planning and Control

15. Manage additional uncertainty. In recovery situations only a part of the flow is valuable, but it is hard to say beforehand which part. This means that sorting and initial testing should be decentralized to separate junk from valuable returns. The same goes for sorting and volume reduction in e.g. plastics recycling. Intrinsic to the push-pull nature of reverse channels, there will often be a mismatch between supply and demand for recyclable products and choice of the right recovery

channels, even in situations with perfect information. E-market places provide a good support tool.
16. Enhance design for recycling. Regarding the environmentally driven network design, in (Tsoulfas et al., 2000), a sector analysis of batteries from the point of view of closed loop supply chains is presented, where several network design criteria are discussed. Decisions to be taken concern modularity, kind of materials, involvement of suppliers (co-design), disassemblability, life cycle considerations (will it last for a long period or a short one?), type of equipment used and standardization of modules/components in the product. Parameters affecting the decision include pollution generated, energy use, residual waste, life cycle cost, production technology, secondary materials, by-products, recyclability, product complexity, product function, and so on.

D.Network structure
17. Match network design with recovery option. Regarding cost and service driven network design, (Fleischmann et. al, 2000) and (Bloemhof-Ruwaard et. al, 1999) give an overview of case studies. They conclude that compared to traditional forward logistics, closed loop supply chains have some distinguishing common characteristics, in particular in terms of processes to be carried out. Typical characteristics of product recovery networks include a convergent part concerned with collection and transportation from a disposer market to recovery facilities, a divergent part for distribution to a re-use market, and an intermediate part related with the recovery processing steps required. Moreover, they derive typical types of networks per recovery option, where they distinguish networks for material recycling, remanufacturing, reusable components, reusable packaging, warranty and commercial returns. These network types generally differ in terms of network topology, the role of and cooperation between actors and the collection and routing system used. Thus, environmental aspects may influence network topology, the role and cooperation between actors and the collection and routing system used. It has been suggested that e.g. facilities should be located close to possible end-users. Such a policy would ease the direct delivery of used products from end-users (Angell and Klassen, 1999).
18. Set up good collections systems. In (Krikke et. al, 2000) a multicriteria model is presented that optimises the supply chain of refrigerators on both economic and environmental (LCA based) criteria. The model is run for different scenarios using different parameter settings such as centralized versus decentralized operations, alternative product designs, varying recovery feasibility and return quantity, and potential EU legislation. The most important conclusion of the project is that, next to efficient logistics combined with optimal product design, system optimality depends on return quality and rate of return. In fact, in this case study these effects outperform the impact of product design and logistics network structure.

Table 1 provides the extended set of design principles, both the traditional and new principles. It also presents the drivers of the new design principles. Note that we do not investigate the actual outcome of applying these design principles on

e.g. network structure. Examples on this can be found in (Bloemhof et.al, 1999) and (Fleischmann et. al, 2000).

It appears that a common denominator is the *life cycle approach*. Both from a cost, customer service and environmental perspective, closed loop supply chains should take the product life cycle –from cradle to grave- as a starting point. From there, supply chain processes, including R&D activities, can be optimised.

Table 1. Extended design principles in closed loop supply chains

Areas	Traditional Principles	What's new?	Consequence	New principles
Organisational	1 2 3 4	optimise on sustainability system complexity uncertainty push-pull convergent network new markets	paradigm shift to product life cycle management, supply chain processes to be optimised from this starting point.	11 12 13
Information	5			14
P&C	6 7 8 9			15 16
Network structure	10			17 18

3 Remanufacturing of Printed Wiring Assemblies at Honeywell

3.1 Case Description

Honeywell Industrial Automation and Control (IAC) is a global player in industrial automation. It produces, supplies and maintains Distributed Control Systems, i.e., both hardware and software to measure, monitor and control production processes of its customers. Distributed Control Systems are networks of intelligent (automated) stations, which control an industrial (chemical) plant or process, where the network distributes logic control, data access and process management. Because of the high capital value of the plants involved and their dependence on control systems, the control systems are often redundant, and need fast service in case of failure. In this case, we study the repair of Printed Wiring Assemblies (PWAs). These PWAs are a critical and valuable component in the TDC-3000 system. They are serviced by well trained service engineers of the national affiliates and regularly replaced due to failure or potential failure. Service contracts oblige Honeywell to respond to a customer call within 24 or 48 hours.

The service process goes as follows. After a call, a service engineer is dispatched to the customer location. The engineer replaces the bad or suspicious PWAs and brings it back in his car to the affiliate depot. Here, the PWA is visually inspected. Some parts are rejected and scrapped, others are tested. Good parts are restocked, either at the depot or the engineer's car. Malfunctioning parts

are returned for repair after authorization of the central logistics department ISLC. ISLC controls all logistics and tactical support for European customers. Local repair by affiliates also takes place, but is not officially approved. PWAs are returned by truck to the central warehouse for Europe in Amsterdam, operated by Van Ommeren Intexo (VOI) and controlled by ISLC. VOI is a logistics service provider that operates the warehouse and transportation for Honeywell. Returned PWAs are consolidated at this central warehouse and from there transported in large batches to the Honeywell production and repair sites in Phoenix (USA) and Johannesburg (SA) by plane (Burlington Air). Here returned PWAs are inspected and, if feasible, repaired or upgraded. Johannesburg is given priority, because it is more dedicated to repair. However, also here regular production takes place. Johannesburg sends repaired PWAs to the central European warehouse and Phoenix restocks them in their own facility for possible supply to Europe or other warehouses worldwide. As soon as a PWA has been replaced and returned, the inventories are replenished, i.e., the affiliate replenishes the engineer's car, the central warehouse replenishes the affiliate depot and the production factories replenish the central warehouse. Return of repaired PWAs from Johannesburg to the European warehouse runs independently from the replenishment procedure. Also in replenishment, inter-continental transportation is covered by plane, intra continental transport by truck.

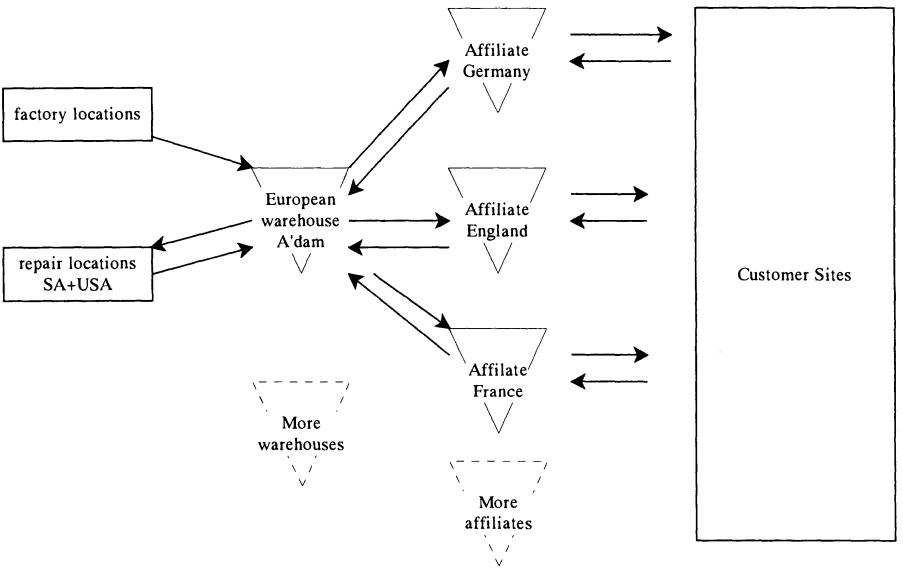


Fig. 2. Honeywell European supply chain for service of TDC 3000-PWAs

Figure 2 represents the existing closed loop supply chain for service of PWAs for TDC-3000 system in Europe, including return flows for repair. Replenishment

rush orders can skip one or more echelons, depending on the cause and location of demand: PWAs can be delivered straight from the factories to the affiliate, from the central European warehouse to the affiliate or from the central warehouse to the customer site. This is done by DHL. Inventories are kept at the production and repair sites, the central warehouse, affiliate locations and in the service engineer's car. In terms of volumes, about 500 are returned each year, of which the majority is remanufactured (note that defects are generally not returned). About 3000 pieces are in stock. Total reverse chain cost are about 360 EURO per item, whereas the reuse value is estimated 700 EURO. The lead-time from the moment of return at the customer site, through repair and back to serviceable stock is approximately 34 weeks, which is a major concern for the management because of the risk of obsolescence. In comparison, the lead-time from production to stock for new spare parts is only 3-4 weeks. For an extensive description of the case, we refer to (Breunesse, 1997). The following shortcomings have been identified:

(i) Return procedure. Non-reusable and reusable PWAs are not separated correctly at the source, because there is no clear procedure and responsibility for return shipping is not assigned. There is a lack of awareness of service engineers that broken PWAs should be returned in a correct manner (e.g. packaged correctly). As a result too many PWAs which are not reusable (due to lack of demand or technical failure) are returned, while reusable PWAs are sometimes not returned.

(ii) Lead-time. Return shipments are done via the forward logistic systems leading to long and also stochastic lead times of returns. In reverse distribution, consolidation times for small return volumes are long and the reusable PWAs that are returned stay too long in the pipeline, due to capacity problems at the production/repair sites. Long lead times are costly due to high capital value of PWAs and danger for obsolescence.

(iii) Availability planning. Also as a result of the long and stochastic lead-time, returned PWAs are not taken into account in the availability planning of ISLC for the VOI warehouse. This is no problem for Phoenix repairables, since they are restocked in Phoenix. However, the Johannesburg repairables are restocked at the VOI warehouse in Amsterdam. They cannot be taken into account in the availability planning, because a lack of logistic control makes lead times long and stochastic, hence Johannesburg is not sufficiently reliable as an internal supplier of PWAs. This is complicated by the fact that valuable information is lost in the long reverse chain, because information does not stay with the product nor is an information system in place to deal with this.

The supply chain improvements suggested in this study were based on the following considerations:

- lead time reduction
- controlling out-of-pocket costs
- no defect PWAs should be returned
- all good PWAs should indeed be returned quickly
- information should be kept with the PWA.

The possibilities for re-design are limited, because existing repair centres (in Phoenix and Johannesburg) cannot be closed down nor can new ones be opened. This is due to the fact that in the near future Honeywell will outsource repair activities. Also, Johannesburg is a priori chosen as the repair facility for Europe, in order to equalize capacity loads company-wide. In other words, only the collection system, goods flows and make-or-buy decisions may be affected. Improvements must be found in simplifying the reverse logistic system. The following solution is suggested: (i) an improved return shipping procedure for service engineers and clients makes sure the right PWAs are returned, (ii) direct shipping by DHL from affiliates to the repair centre in Johannesburg reduces lead time and makes it easier to keep information with the PWA. The forward system and the shipping from Johannesburg back to stock in central European warehouse remain the same. This reverse supply chain has a better performance, while unit out-of-pocket cost increases from 360 EURO to 373 EURO per PWA. However, total shipping costs may be reduced because useless returns are avoided by the improved shipping procedure.

3.2 Analysis

In the old situation the supply chain is not geared for return flows. It appears that Honeywell's design principles were limited to:

- Use the forward supply chain as much as possible
- Minimize out-of-pocket costs per stage in the reverse chain.

The management of Honeywell recognised that lousy performance necessitates the re-design of the supply chain for repair of PWAs. To this end a subset of design principles presented in Section 2 is applied. Table 2 gives an overview of the design principles and their application in the re-design. They are explained below.

Table 2. Overview of design principles application at Honeywell

principle	description	yes, no, somewhat
1 (A)	link to business strategy	somewhat
2	exploit human resources	no
3	form strategic alliances	no
4	target service levels	yes
5 (B)	power of information	yes
6 (C)	focus on financial performance	somewhat
7	manage comprehensively	somewhat
8	manage details	no
9	performance mgt.	somewhat
10 (D)	leverage volumes	yes
	EXTENDED	
11 (A)	**use sustainable suppliers**	**no**
12	**create new markets**	**no**

Table 2. Overview of design principles application at Honeywell (continued)

principle	description	yes, no, somewhat
13 (B)	**use new management tools**	**no**
14	**use life cycle accounting systems**	**no**
15 (C)	**manage uncertainty**	**somewhat**
16	**enhance design for recycling**	**no**
17 (D)	**match network and recovery option**	**yes**
18	**enhance rate and quality of return**	**yes**

A: organisational, B: IT, C: planning and control, D: network structure

Honeywell should consider closed loop supply chains management as an essential part of their service functions because returns are often the only source for spare parts, especially for phased out products. Using a carrier for speeding up returns reduces lead-time and variance and thus risk of obsolescence and uncertainty. By outsourcing, the network structure is automatically changed. It also avoids organizational and responsibility problems and leverages volumes. However, this is also quite costly, in particular with long distances from Europe to South Africa. A centralized network in Europe would be more feasible, however the phasing out of proprietary hardware of Honeywell and thus the phasing out of own repair operations makes this infeasible. Here we see that business strategy, i.e., the decision to use IBM hardware and to phase out proprietary hardware, has an impact on closed loop supply chains decisions. Decentralized testing and a simplified returns procedure aim at improvement of rate and quality of return. Keeping the information with the PWAs clearly enhances the power of information. The study described by (Breunesse, 1997) is an extensive performance measurement, however follow up is unclear. The definition of cost is widened, now including obsolescence cost.

Taking a closer look, we see that Honeywell has focused on logistics, operations and information aspects in order to optimise costs and availability. But did Honeywell do the right thing from an environmental point of view? The use of airplanes over long distance is an energy consuming activity. Although reusable PWAs are well taken care of, non reusables are scrapped by local firms, of which no information is available. The company appears to have little 'product life cycle consciousness'. No sustainable suppliers are used. Moreover, product modularity or more general product design aspects have never been nor will be an issue. The company is unaware of future environmental legislation on producer responsibility, although at the time this was being prepared by the European Union. The phase out of propriety hardware might help Honeywell in this respect, however this is pure coincidence. In conclusion, sustainability remains a subordinate issue.

4 Discussion and Conclusions

Traditional wisdom holds that sustainability is costly and the domain of environmental idealists. Few companies have established closed loop supply chains and the ones that have usually implemented end-of-pipe solutions are enforced by law. (Stock et. al, 1998) conclude –amongst other things- that "the state of development of Reverse Logistics is analogous to that of inbound logistics of 10-20 years ago". For example, in the reverse chain, next to out-of-pocket costs we must also include obsolescence costs and service related criteria. This is in fact a very old principle. The importance of lead time effects both on costs and service level has been extensively reported in classic logistics literature. However, our hypothesis is that most business companies will –like Honeywell- reinvent the wheel at this point.

With regards to the extended design principles it appears that the adoption of the product life cycle paradigm is the basis of all. It is necessary to develop and design recoverable products, which should be technically durable, repeatedly usable, harmlessly recoverable after use and environmentally compatible in disposal (Gotzel, et. al, 1999). However, optimisation of supply chain processes can add significant results. Extending service and enhancing function, especially at the usage phase, improves eco-efficiency and reusability (Tsoulfas and Pappis, 2001). Modularity and standardization also improves opportunities for repair and (cross-supply chain) reuse of components and materials. Suppliers that are sustainable should be selected and involved in product design and component repair. Thus, reverse chains should be designed taking the product life cycle as a starting point, but a (partial) redesign of the forward chain may be required as well. A number of management tools, such as environmental assessment, life cycle analysis, environmental accounting methods, but also 'simple' logistics principles can help business identify and select opportunities for improvement.

In conclusion, closed loop supply chains are fundamentally different from traditional 'one-way' supply chains, particularly in view of sustainability. As a result, traditional design principles need to be extended. However, also traditional principles apply to closed loop supply chains. We suspect that both are often not well understood by business practice. Obvious improvements can be made by applying traditional principles. This is the easy part. New principles are necessary to reduce emission and waste, and require life cycle driven approaches supported by advanced management tools such as LCA and LCC. A new attitude is needed, both with supply chain actors and consumers.

References

Angell, L.C. / Klassen, R.D (1999): Integrating environmental issues into the mainstream: an agenda for research in operations management. *Journal of Operations Management*, 17:575-59

Bloemhof-Ruwaard, J.M. / Fleischmann, M. / van Nunen, J.A.E.E. (1999): Reviewing distribution issues in reverse logistics. Lecture Notes in Economical and Mathematical Systems, 480, pp. 23-44, Springer Verlag, Berlin, Germany, edited by. M. G. Speranza and P. Staehly

Breunesse, H. (1997): *Control in reverse logistics.* Masters Thesis, Twente University, Mechanical Engineering, Enschede, The Netherlands

Caldwell, Bruce (1999): Reverse Logistics. *Information week online*, April issue, www.informationweek.com/729/logistics.htm

Corbett, C.J. / Van Wassenhove, L.N. (1993): The Green Fee: Internalizing and Operationalizing Environmental Issues. *California Management Review*, 36 (1):116-135

Faucheux, S. / Nicolaï, I. (1998): Environmental technological change and governance in sustainable development policy. *Ecological Economics*, 27: 243–256

Ferrer, G. (1997): Communicating developments in product recovery. Working Paper 97/30/TM, INSEAD, France

Fleischmann, M. / Krikke, H. R. / Dekker, R. / Flapper, S. D. P. (2000): A characterisation of logistics networks for product recovery. *Omega*, 28 (6):653-666

Fleischmann, M. (2000): *Quantitative models for reverse logistics.* Ph.D. thesis, Erasmus University, Rotterdam, The Netherlands

Gattorna, J.L. (1997): *The Gower Handbook of Logistics and Distribution Management.* Gower, Vermont, USA

Gotzel C. / Weidling, J.G. / Heisig, G. / Inderfurth, K. (1999): Product return and recovery concepts of companies in Germany. Preprint Nr. 31, Otto-von-Guericke University of Magdeburg, Germany

Klassen, R.D. / Angell, L.C. (1998): An international comparison of environmental management in operations: the impact of manufacturing flexibility in the U.S. and Germany. *Journal of Operations Management*, 16 (3–4): 177–194

Krikke, H.R. / Bloemhof-Ruwaard, J. M. / Van Wassenhove, L.N. (2000): Design of closed loop supply chains for refrigerators. Working Paper, Erasmus University, Rotterdam, The Netherlands

Ralph Sims, E. (1991): *Planning and Managing Industrial Logistics Systems.* Elsevier, Amsterdam, The Netherlands

Stock, J. (1998): Reverse Logistics Programs. Council of Logistics Management, USA

Tsoulfas, G.T. / Pappis, C.P. / Minner, S. (2000): A sector analysis of batteries: the perspective of reverse logistics. REVLOG Summer Workshop, Lutherstadt Wittenberg, Germany

Tsoulfas, G.T. / Pappis, C.P. (2001): Application of environmental principles to reverse supply chains. In proceedings of the 3rd Aegean Conference, Tinos, Greece, May 19-22

A Behavioral Approach for Logistics System Analysis and Design: A Reverse Logistics Case

Marco Mazzarino[1], Raffaele Pesenti[2], and Walter Ukovich[3]

[1] Dipartimento di Pianificazione
IUAV - Istituto Universitario di Architettura di Venezia
c/o ISTIEE - Università degli Studi di Trieste
[2] Dipartimento di Ingegneria Automatica ed Informatica (DIAI)
Università degli Studi di Palermo
[3] Dipartimento di Elettrotecnica, Elettronica ed Informatica (DEEI)
Università degli Studi di Trieste

Abstract Traditional logistic system analysis quite often assumes a single decision-maker (the planner) operating in a state of complete information and full decision power. He pursues the objective of designing an efficient logistic network by solving a sequence of operational problems mainly in the form of optimization models. More realistically, one should consider that the decision power is actually distributed within the logistics system among different actors (agents or holons) having different (conflictual or cooperative) goals, following different behavioural rules and generating interdipendencies. The shift from a SAS (single-agent system) approach to a MAS (multi-agent system) one induces significant changes in the techniques and models adopted for studying the logistics systems. In particular, one has to take into account the impact of the different coordination mechanisms (stemming from the coordination theory) onto the decision-making process of each agent (holon). We develop a MAS approach for a reverse logistics case related to the urban waste and present some preliminary results in terms of logistic system design and modelling.

1 Introduction: The Classical Approach (SAS - *Single-Agent System*) to Logistics System Analysis and Design

The traditional approach to logistics system analysis and design refers to classical planning (see Bianco (1987)). It identifies on a geographic scale a number of logistics activities (transportation, collection, etc.) generating a number of logistics problems. Models are then develop in order to solve these problems coming up with a final "optimal" configuration of the logistics system (largely in terms of a minimum total cost).
The most important problems are the following ones:

- location problems:
  - where logistics facilities should be located and in which number;
  - how markets should be efficiently allocated to facilities;
- transportation problems:

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

- which transport service or mode should be chosen;
- how recovered products and materials should be efficiently loaded on vehicles;
- what should be the composition and dimension of the fleet;
- when transport activities should be performed and which routes should be chosen.

The most important models developed for this kind of problems are:

- location models such as the Uncapacitated Facility Plant Location and the Capacitated Facility Plant Location ones;
- location-allocation models;
- mode choice models;
- loading models;
- fleet size and composition models;
- vehicle routing and scheduling models with and without time-windows.

The more or less implicit assumption on which such an approach is based upon is that a single decision-maker is identified (the "planner") having full decision power and complete information over the most relevant variables of the logistics system. More specifically, it is assumed that some variables (decision variables) are completely under control of the single decision-maker while the rest of the variables are taken as given by him (parameters), that is, they must be provided to the single decision-maker. We can referred to this as the case of *centralized decision power* and *symmetric information.*
By managing the decision variables and parameters the single decision-maker must take decisions about how to efficiently use resources in order to come up with an optimal configuration of the logistics system. Therefore, he solves a number of logistics problem at the optimum. The contribution of the theory is that of providing the decision-maker with a number of models helping him to solve the logistics problems. Usually he *sequentially* solves a number of optimization problems, from the strategic to the operational ones.
Conversely, if one assumes more realistically that within logistics systems (and generally speaking, in every economic system) the decision power and the information over the most relevant variables are actually *distributed* among a number of different economic agents, as the behavioural approach does, it is no longer possible to approach the issue of the logistics system design in terms of a sequence of optimization problems to be solved by a single decision-maker. Indeed, the reasoning can no longer be referred to the "system" in aggregate, rather to each economic agent.
In other words, we identify the following shortcomings of the classical approach to logistics system planning and design:

- little, if any, attention paid to the identification of the different economic agents within a logistics system and to their behavioural characteristics;
- as a consequence, lack of analysis of the main interrelationships among agents and, above all, of the their impact on the economic behaviour and decision processes of the agents;

- as a further consequence, no emphasis posed onto the allocation problem, that is, on how welfare is allocated among agents.

The main purpose of the paper is to propose a new approach for analysing and designing logistics system in a Supply Chain Management context, that is, a behavioral one and compare it with the classical ones. In particular this new approach is applied to a reverse logistics special case.
In particular it is worth to note the possible contribution that the economic theory can give in this framework:

- from a *positive* standpoint, in order to understand how sustainable economic and logistics systems really work, and to evaluate their performance;
- from a planning and policy standpoint (*normative*), in order to properly manage the systems so as to maximise their efficiency (in a paretian sense). In this manner, the approach could be seen as a Decision Support System for decision-makers having planning and policy objectives.

The paper is organized as follows. In the next section it is highlighted how logistics systems are part of the more general economic system. In particular, it is stressed that the concept of economic system should be viewed in terms of *sustainable* economic system.
On the basis of the deficiences of the classical approach (that we name SAS - single agent system) on Secction 3 we develop some ideas for a *behavioural* approach which is based on the concept of *holonic* system. We referred these ideas to as a MAS (multi-agent system) approach. Section 4 presents the modelling development of this approach which ends up with a general modelling framework. Then, a case-study is proposed with reference to the urban waste logistics system in Italy and specifically in Friuli-Venezia Giulia region. The case-study develops through an empirical analysis of the holonic system and it comes up with a specification of the general modelling framework.
Finally, some conclusions are drawn above all in terms of comparison between the SAS and MAS approach and some future lines of research are indicated.

2 The Logistics and Sustainable Economic System

Logistics systems are part of the economic system. Since recently, the interpretation of the concept of economic system has been enlarged so as to comprise the sustainability issues, mainly as a consequence of the negative environmental effects produced by the economic growth process (see Boulding (1966)). Such an issue becomes relevant when one considers the interrelationships between the economic and the environmental system in enabling both systems to co-exist and reach an equilibrium (see Pearce et al. (1990)). It should be noted that the two systems are separately dealt with by economists and natural scientists respectively. Natural scientists are mainly concerned

with issues such as, for instance, how ecosystems survive, while economists had produced lots of efforts in order to find an *internal* economic equilibrium (see Arrow et al. (1954)). The interrelationships between the two systems and how they can be managed is a quite recent field of research (see Daly (1980)).
One of the most famous model of a sustainable economic system is the so-called circular (or closed) economic system model developed by Pearce and Turner (1990) (see figure 1). It is based on the traditional neoclassic model, ie the *linear* one, with the addition of the interrelationships between the economy and the environment. The model highlights, among other things, two fundamental economic functions of the environment:

- the environment providing input resources to the economic system;
- the environment receiving waste products and materials from the economic system (on the basis of its "carrying capacity").

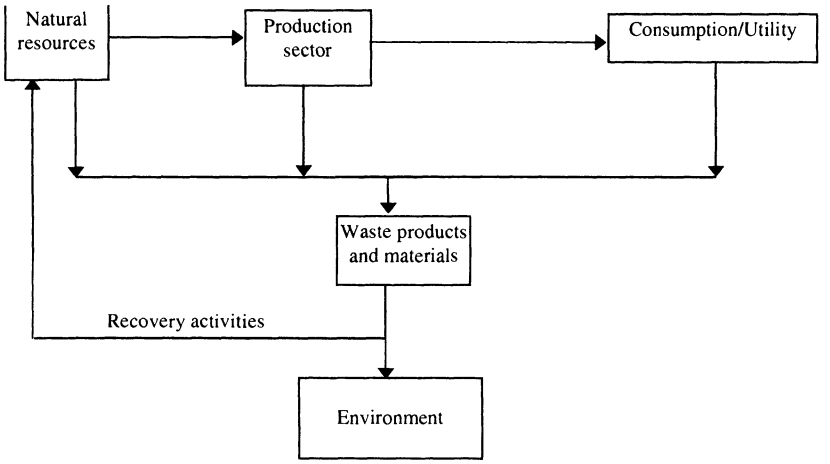


Fig. 1. The sustainable economic system

Moreover, a general link exists between the input flows (resources) and the output flows (waste products) of an economic system and it is based on the thermodynamics laws. A number of economists have put emphasis on these aspects (see Daly (1986) and Georgescu-Roegen (1971). On the basis of the interrelationships between the economy and the environment some *sustainability rules* are put forward in terms of rules which the economic system must follow in order to guarantee its sustainability. These rules consist in a number of constrains posed on the level of exploitation of natural resources (first rule) and on the level of waste products and material disposed into the

environment (second rule).
Within such a sustainability framework one of the fundamental kind of activity for the sustainability goal to be achieved is that of *recovery*, that is, all types of activities enabling the recycling of waste products by re-transforming them into natural resources capable of being used by the economic system. More specifically, recovery activities can be classified in re-use, remanufacturing and recycling ones (see Fleishmann et al. (1997)). It can be shown that recovery activities help an economic system to increase its sustainability level by:

- reducing the exploitation level of natural resources;
- reducing the negative impact on the carrying capacity of the environment.

2.1 The Logistics Perspective

The sustainable economic system can be represented in a more articulated and actually complete manner if one highlights its logistics characteristics. In fact, the flows of goods and materials moving from the production sector to the consumption one, and vice versa, are managed within logistics systems. Such systems are nevertheless neglected by mainstream economics in so far as only two agents are mainly considered, i.e., industrial firms and households. The management and organization of flows going from producers to consumers is the field of the so-called *forward* logistics, while flows from consumer back to the production sector are studied by the *reverse* logistics. The latter is a recent field of research and it seems to be quite intriguing both for practitioners and theorists (see Ayres (1997), (1998), Barros (1998), Bloemhof-Ruwaard et al. (1999), Ferrer (1996a), (1996b), (1996c), (1997), Ferrer et al. (1998), Fleishmann et al. (1997), Johnson (1998), Kroon et al. (1995), Pohlen et al. (1992), Rose (1994), Thierry et al. (1995), Van der Laan (1998)).
In order to have an illustration of the logistics characteristics of a sustainable economic system one can refer to figure (see 2). More specifically, we define the *sustainable logistics* as *the field of research studying the efficiency of management of both the forward and the reverse flows in an integrated way in order to give a contribution to the sustainability issue.*

It is quite clear that the perspective here adopted is that of Supply Chain Management (see Ballou (1999), Bechtel et al. (1997), Cooper et al. (1997), Motwani et al. (1998)). In particular, we stress the importance of recovery activities at system-level (rather than at a corporate level) as a decisive factor contributing to the sustainability of an economic system. The analysis of a sustainable economic system in which the logistics aspects are highlighted is the subject of this paper. The main purpose consists in proposing a quite innovative approach for analysing and designing such systems, that is, a behavioral one vs. the classical one. In the end, we come up with a general modelling framework.

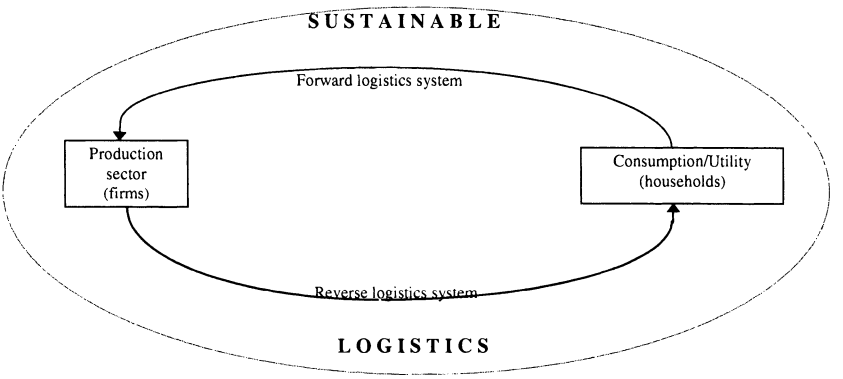


Fig. 2. sustainable economic system: logistics charcaters

3 Towards a Behavioural Approach: The MAS (*Multi-Agent System*) Perspective. First Modelling Results

In order to tackle the deficiencies of the SAS approach, it should be firstly noted that in every real economic system the decision power and the information are distributed among a number of agents developing a series of interrelationships. Reality shows us *multi-agent systems*, that is, a number of agents performing their activities in a dynamic context by coordinating their decisions in a cooperative and/or conflictual manner. We can referred to this as the case of *distributed decision power* and *asymmetric information* (see Schneeweis(1999)). Every agent therefore has a certain "amount" of decision power and information over some variables of the system. The behavioral approach is then based on a *holonic* structure of the system (see van Brussel et al. (1997a), 1997b)). A *holon* is an entity having two characteristics:

- autonomy;
- cooperation/coordination

Based on the first characteristic each *holon* (agent) take some optimal decisions in an independently manner with respect to the other agents. Nevertheless, the second characteristic shows how the behavior of each holon is strategically "linked" with that of the other holons. The behavioral approach puts emphasis on these *holonic* characteristics of an economic and logistics system, in particular raising the following questions:

- For the *autonomy* characteristic (the so-called *local conditions*):
  - which are the economic agents of the system and which are their economic behaviours and decision processes (objectives, kind of decisions

to be taken, kind of problems to be tackled, variables involved, etc.). These questions stress the need for a strong *empirical* analysis to be performed in so far as the local conditions should be studied "on the field";

- For the cooperation/coordination characteristic:
  - what kind of interdependencies exist among agents;
  - how they can be represented;
  - which are the *coordination mechanisms* managing such interdependencies.

3.1 The Coordination Theory

The need of identifying the interrelationships among agents in order to define their economic nature and, above all, the need of understanding their impact on the decision processes of the agents represent one of the most crucial aspect of the behavioral approach. We can referred to the so-called *coordination theory* (see Malone (1987), (1988), (1992), Malone et al. (1991), (1994)), which is a quite recent interdisciplinary field of research encompassing the decision theory, computing science (see Malone (1992)), linguistics, economics, game theory, organization theory (see Malone et al. (1991), (1993)), etc.
What does *coordination* mean? A possible definition is: *managing dependencies among different activities* (see Malone et al. (1991)). In fact, it is quite clear that if there is no interdependency among different activities performed by different agents there is nothing to coordinate. The most relevant issue is therefore that of identifying and classifying the different types of interdependencies so as to define the different categories of coordination mechanisms. In the main, given a number of interrelationships among a number of agents we can define a coordination mechanism as the *set of rules enabling some activities to be assigned to the agents.* The concept of coordination is at the core of economics. One simple and famous example of interdependencies among different agents who coordinate their activities in order to achieve a common objective is given by the competitive market mechanism. This is simply the case of a decentralized coordination mechanism aiming at an optimal allocation of resources (the so-called "invisible hand") (see Baligh (1986), Kurose et al. (1989), Lumer et al. (1990), Miao et al. (1992), van Prunak et al. (1997a), (1997b)). The planning process described by the classical approach can be seen as a strongly centralized coordination process in which each agent sends information regarding the system variables to the single decision-maker.

It should also be pointed out that the classical approach does not raise the issue of the welfare allocation among agents. This is quite clear given that just *one* decision-maker is considered. What it is relevant for the classical approach is the general performance *of the system* without considering the individual performances of the agents and the relationships between individual performances and the system performance in terms of welfare.

From a behavioral point of view, the main goal is to build a model of the sustainable logistics system capable of incorporating the local conditions, the strategic interrelationships of the agents and the allocation problem. The main contribution that the theory can give in this context is that of developing a number of behavioral models referred to each economic agent incorporating their interdependencies while trying to solve, at the same time, the allocation problem.

3.2 The General Framework Model

Modelling a sustainable economic and logistic system can be done by defining the following elements (see Sikora et al. (1998)):

- the *individual performance function* for each agent, with regard to the "autonomy" characteristic of each holon;
- the *interaction variables*, with regard to the interrelationships existing among agents and the respective coordination mechanism (characteristic of "cooperation/coordination");
- the *global performance function*, with regard to the allocation problem.

The general model can be expressed as follows:

$$\pi_1 = \pi_1[\bar{P}_1(...., \bar{z_{ij}}, ...)] \quad (1a)$$

$$\pi_2 = \pi_2[\bar{P}_2(...., \bar{z_{ij}}, ...)] \quad (1b)$$

$$....$$

$$\pi_n = \pi_n[\bar{P}_n(...., \bar{z_{ij}}, ...)] \quad (1c)$$

$$....$$

$$z_{ij} = ...$$

$$....$$

$$\Pi = \phi(\pi_1, \pi_2, ..., \pi_n) \quad (1d)$$

where:
π_i = individual performance
$\bar{P}_i$ = decision problem sets
z_{ij} = interaction variables
Π = index of global (system) performance
ϕ = global performance function

In the model, the individual performance of each agent i, π_i (i=1,..,n), depends on some decision problems - $\bar{P}_i$(i=1,...,N)- in which the so-called interaction variables - z_{ij} - are considered.
Generally speaking, the most relevant variables related to each agent (and therefore to his performance) can be divided in three main categories:

- decision variables;
- parameters;
- interaction variables.

This classification stems from the fact that the decision power and the information are distributed within the system among agents. Each agent therefore has:

- full decision power over some variables, that is, he can independently and fully decided about the characteristics of some variables; these are the so-called decision variables;
- no decision power over other variables, that is, variable characteristics are fully determined by other agents and the agent has to take them as given; these are the so-called parameters;
- partial decision power over some other variables. That means that he has some influence on these variables but they are fully determined by the interaction between he and another agent (or more than one) by means of coordination mechanisms; these are the so-called interaction variables.

It is quite clear that we can identify as parameters in the performance function of an agent:

- decision variables of other agents;
- interaction variables defined by other (different) agents.

In other words, the proposed classification is valid relatively to each agent.

It should be noted that the interaction variables may not necessarily be expressed by equations, rather they can come out from a number of submodels representing the coordination mechanisms.
Another crucial point of the general framework model is that of the global performance function. While the classical approach expresses the optimal configuration of the logistics system usually in terms of a minimum total cost, the behavioral approach represents it in terms of the Π value. The objective is that of finding the optimal welfare allocation among agents: given the ϕ function one has to find the optimum vector $(\pi_1, \pi_2, ..., \pi_n)$ (in a paretian sense), that is, that vector maximising the Π value. More generally, one can think of building an *efficiency frontier*, i.e., all the optimal values of Π in a paretian sense. The critical point for such values to be found are the characteristics of the ϕ function. In other words, one has to find out how the individual performances are linked with the performance of the system. Basically, two simple expressions can be identified (see Sikora et al. (1998)):

- an additive one, that is, $\Pi = \sum(\pi_1, \pi_2, ..., \pi_n)$;
- a competitive one, that is, $\Pi = max(\pi_1, \pi_2, ..., \pi_n)$.

However, the specific form of the ϕ function can be rather complex. This certainly represents a future line of research.

4 A Case-Study: The Urban Waste Problem

We can summarize the step-by-step methodology used by the behavioral approach as follows:

- Empirical analysis:
  - identify the agents;
  - qualitatively describe their behaviors and decision processes and their interrelationships;
- Modelling analysis:
  - identify the main variables of the system;
  - assign them to each agent by classifying them into:
    - decision variables
    - parameters
    - interaction variables
  - model the economic behaviors and the interrelationships;
  - develop and specify the general model.

Now we try to apply the general methodology of the behavioral approach to a specific situation: the reverse logistics system of the urban waste in Italy. As a matter of fact, sustainable logistics implies applying the general methodology of the behavioral approach to both the forward and the reverse flows in an integrated way. However, trying to consider the whole system means that one should take into account all types of forward-reverse flows and all the agents in the system.

For now, we focus on a specific reverse system which is, as one will see, rather complex for itself, assuming for simplicity no interrelationships between forward and reverse chains.

To introduce the issue, let us stress *disposal* activities will no longer exist in the medium-long run in industrialized countries due to legislation in many countries that provide for a maximisation of recovery activities. Generally speaking, the recovery activities can be of two types:

- material recovery;
- thermal recovery.

In Italy the law distinguishes between "special" waste and "urban" waste. The former refers to waste having mostly an industrial origin, while the latter indicates waste largely produced by households. As said, we concentrate on the second type of waste for which Italian laws foresee the following policy scenario to be reached in a few years:

- a significant increase of the so-called "separated collection", whose levels are currently too low;
- the remaining quantities and types of urban waste must be incinerated, therefore recovering energy;

- only the sub-products of the incinerated materials must be landfilled, that is, no urban waste must directly be landfilled.

The types of different materials and sub-products (and therefore the different reverse logistics chains) which can be obtained by the separated collection are left to local authorities. In Friuli-Venezia Giulia, the minimum policy scenario specifies three categories of materials to be recovered by separated collection:

- "dry" recycling materials (plastic, paper, metals);
- glass;
- organic waste.

Having chosen the waste logistics system we then apply the general methodology by:

- identifying the main agents of the system;
- describing their economic behaviors, decision processes and interrelationships by means of empirical analyses (mainly, by direct interviews);
- identifying the main variables of the system;
- assigning them to each agent and classifying them into decision and interaction variables and parameters. We concentrate then onto the interaction variables and identify the main coordination mechanisms;
- describing the economic behaviour of the agents in modelling terms based upon the variables assigned to them ;
- detailing the general framework model.

4.1 Identification of the Main Holons of the Urban Waste Reverse Logistics System

The first step in applying the behavioral approach is the identification of the main agents (public and private) consituting the (general) reverse logistics system of the urban waste. In simplified terms, they are:

- market demand holons;
- market supply holons;
- logistics system holons in a strict sense.

The market demand holons are represented by firms using recovered materials, while market supply holons are households generating urban waste.
The logistics system holons in a strict sense are:

- holons on the strategic level, mainly public bodies (State, local authorities, counties, etc.);
- holons on the operational level, that in turn are given by:
  - operators managing logistics services;
  - operators managing logistics facilities.

Operators managing logistics services are those involved in the collection and transportation of the urban waste, while the operators managing the facilities are given by:

- disposal plant operators;
- incinerator operators,
- operators of testing, reprocessing, composting, sorting plants.

We can also have the so-called integrated operators, that is, operators managing both facilities and collection/transportation activities of various kind of urban waste.
The agents of the strategic level, ie public bodies, follow a planning process aiming at defining on a certaing regional territory:

- the zoning for generation and processing of waste materials;
- criteria for locating the processing and disposal logistics facilities, on the basis of a "self-sufficiency" principle for each zone. Sometimes these operators also decide the specific location;
- the main characteristics of the collection and transportation activities.

The municipal authorities are the links between the operators on the strategic level and those who pratically performed the logistics activities, i.e., the operators of the operational level. In fact, they are charged with the assignment of the collection, transportation and incineration activities to operational agents usually by means of public auctions. For facilities such as those related to material recovery and landfilling a substantial deregulation regime operates. This means that any agent regularly owning a license for performing recovery and disposal activities can do this without any further procedure or requirements. At a more tactical/operational level, public bodies decide the type and amount of incentives and subsidies to be assigned to operational agents in order to promote their activities while reaching some social objectives.
To summarize, it appears quite clear that the planning activities related to the reverse logistics system of the urban waste is implemented through a two-steps process: the strategic decisions with regard to zoning and location problems are taken first on the basis of a public process, while the operational decisions are taken consequently on a second step.
A simplified picture of the holonic system can be seen in figure (see 3).

4.2 Identification of the Main Variables of the System

Having outlined the main behavioral characteristics of the system agents, we represent the system in a simplified way by means of the following agents:

- CTO: collection/transporation operators, numbered with r;
- FO: facility operators, numbered with j;
- PB: public bodies, represented by a single agent;

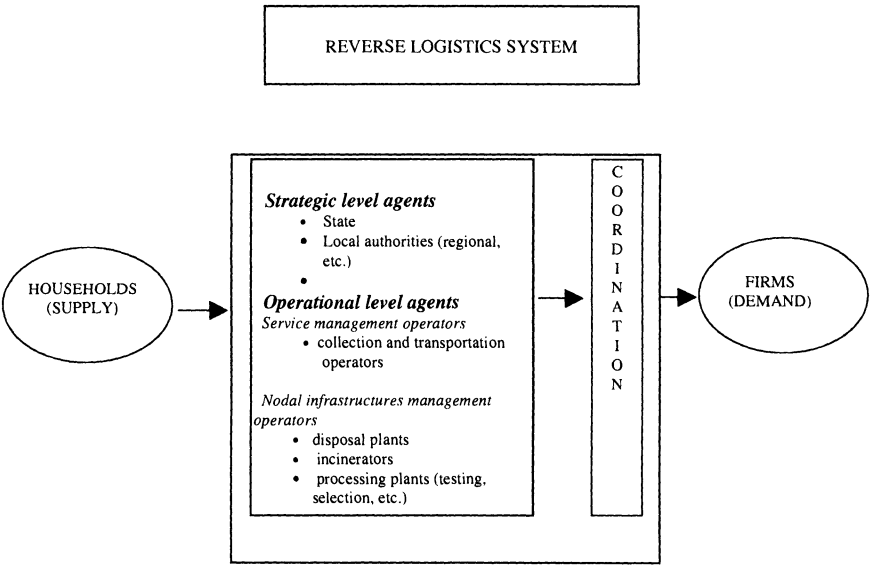


Fig. 3. The holonic system

- H: households, numbered with i;
- M: market firms, numbered with l.

We then identify the following main quantitative variables of the system:

- c_{ijr}: the transportation and collection cost from pick-up point i to point j (processing facility of operator j) in the logistics network for operator r;
- X_{TOT}: total amount of waste produced by households, which is given by the fraction of separated collection (X_{SC}), i.e., the "real" reverse flows, plus the residual (X_R), which is usually incinerated or disposed. Specifically, if $i = 1, ..., n$ is the number of households/collection points we have: $X_{TOT} = \sum_i x_{SC_i} + \sum_i x_{R_i} = X_{SC} + X_R$;
- x_{ijr}: the amount of separated waste produced in point i and transported in j (processing facility) by operator r. We have $\sum_i \sum_j \sum_r = X_{SC}$;
- f_j: investment costs (fixed costs) of a logistic facility operator j;
- I: number of collection points in the network (assumed to be equal to the number of households), $I = 1, ..., n$;
- $\mu_r, \bar{\mu}_j$: expected profit for operator r and j involved in auction activities;
- q_{jl}: amount of recovered products sold on market l by facilities operator j;
- P_r: auction bid by CTO operator r ;
- $\bar{P}_j$: auction bid by FO operator j ;

- p_j: selling price of recovered materials by facility operator j;
- a_{kr}: capacity of vehicle k for the collection and transportation activities of operator r;
- K_r: number of vehicles of operator r;
- ct_j: processing cost for operator j;
- τ: amount of (financial) incentive from public bodies, which can be assigned to investment cost (τ_{f_j}), to collection/transportation cost ($\tau_{c_{ijr}}$), to selling price (τ_{p_j}) or to households (τ_{SC_i}).

We also define a technical input-output coefficient for operator j as:

$$b = \sum_l q_{jl} / \sum_i \sum_r x_{ijr} = \sum_l q_{jl} / X_{SC}$$

It shows how many units of recovered products can be obtained at a facility and sold on the market (output) given a number of units of waste entering the production process (input).
Having identified these variables, we assign them to the three simplified and representative figures of logistics operators, i.e., public bodies (PB), collection/transportation operators (CTO) and facility operators (FO). At the same time we classify the variables assigned to each operator in decision and interaction variables and parameters.

On the basis of the assignment and classification of variables to agents we can describe more in depth the economic behaviors of the agents and try to model them.
Public bodies (PB) have a number of objectives which can be expressed in terms of a social utility function U_s (for instance, they should improve recovery activities, respect some environmental constraints, achieve a self-sufficiency objectives for each zone, etc.). In doing this they use regulation tools such as incentives and taxes. We assume that the amount of incentive τ is an interaction variables whose characteristics depends on the relationships between PB and the agents to whom the incentives are assigned. This means that τ_{f_j}, $\tau_{c_{ijr}}$, τ_{p_j} and τ_{SC_i} are interaction variables for PB. The individual performance of PB can be defined in terms of a level of public spending to be minimized. In fact, PB aims at minimizing public spending (and therefore tax pressure) while reaching some social welfare goals. The social utility function is actually a complex one. We can assume that it depends upon a number of social goals $\epsilon_1, ..., \epsilon_n$ to be achieved in a satisfying way. Therefore we define the individual performance of PB as follows:

$$\begin{aligned} \pi_1 &= \tau_{SC} X_{SC} + \sum_r P_r + \sum_j \bar{P}_j + \tau_{f_j} \sum_j f_j + \\ &+ \tau_{c_{ijr}} \sum_i \sum_j \sum_r x_{ijr} c_{ijr} + \tau_{p_j} \sum_j \sum_l p_j q_{jl} - T X_R \qquad (2a) \\ s.t.\ & U_s(\epsilon_1, ..., \epsilon_n) = \bar{U} \qquad (2b) \end{aligned}$$

The number of collection points I in the network is assumed to be decided by the CTO, even if in reality it is frequently decided by PB by defining the auction rules. Variables P and $\bar{P}$ are considered interaction variables since they come out from public auction mechanisms.
The X_{SC} variable is decided by households on the basis of the amount of incentives to generate separated refuse collection - τ_{SC} - determined by the interaction with PB. T indicates the tax payed by households on non-separated refuse collection (waste to be incinerated or disposed) and it is determined by PB. More specifically, we can assume households having a potential reverse flows supply function such as:

$$x_{SC_i} = f(T, \tau_{SC_i}) \quad i = 1, .., n \tag{3}$$

As for the collection/transportation operator (CTO) we can assume he is a profit-maximising agent. His economic behavior can be summarized in the following individual performance:

$$\begin{aligned} \pi_{2r} &= P_r(\mathcal{A}) - FC_r - VC_r\{K_r, I, a_{kr}, c_{ijr}(\tau_{c_{ijr}}, \mathcal{M}), x_{ijr}[x_{SC_i}(\tau_{SC_i})]\} \\ &= \mu_r(\mathcal{A}) \end{aligned} \tag{4}$$

In the formula, in addition to the already known variables, we indicate $P_r(\mathcal{A})$ as the bid offered by the CTO within the public auction which is dependent on the correspondent coordination mechanism; FC_r are the fixed costs and μ_r is dependent on a coordination mechanism as well. The CTO activity produces some fixed and variable costs, FC_r and VC_r, that should be minimized. Within the variable costs he can decide upon the variables K_r, I and a_{kr}. In other words, he can decide on the number of vehicles to be used, on the number of collection points in the network and on the capacity of the vehicles. Generally he solves some classical problems such VRP, loading, fleet size and composition, etc. in order to optimize his activities.
Assuming CTO is an integrated operator, the c_{ijr} variable is determined by the interaction between the CTO and the "pure" transport operator (to) on the basis of a market mechanism and it is also dependent upon the amount of incentives received by PB. The x_{ijr} variable depends on the x_{SC_i} variable which is decided by households and in turn depends on the amount of incentive received by households from PB (τ_{SC_i}). Finally, the difference between P_r and the total costs gives the expected profit for CTO, μ_r, for which the public auction's rules assures it is a true revealed value (that is, CTO does not behave strategically), therefore achieving efficiency in the allocation of activities (see Myerson (1981)).

The facility operator FO can also be assumed as a profit-maximising agent. Therefore, he aims at maximising his profit in two likely situation:

- if the facility management is assigned by means of a public auction, the FO will be subject to the same constraint regarding μ as the CTO operator;

- if the facility can be managed in a deregulated manner, that is, everyone having a regular license can operate a recovery or a disposal facility, the FO will not be subject to the constraint on μ; in this case he will largely rely upon market revenues.

The economic behaviour of FO can then be modelled, in the two likely situations, as follows. In the first case:

$$\pi_{3j} = R_j[q_{jl}(\mathcal{M}), p_j(\tau_{p_j}, \mathcal{M}), \bar{P}_j(\mathcal{A})] - f_j(\tau_{f_j}) + \\ - VC_j[ct_j, x_{ijr}(\mathcal{CL})] = \bar{\mu}_j(\mathcal{A}) \tag{5}$$

In the second case:

$$\pi'_{3j} = R_j[q_{jl}(\mathcal{M}), p_j(\tau_{p_j}, \mathcal{M})] - f_j(\tau_{f_j}) + \\ - VC_j[ct_j, x_{ijr}(\mathcal{CL})] = \bar{\mu}_j \tag{6}$$

In both cases FO aims at maximising his profit which is given by revenues minus fixed and variable costs. The difference is that in the first case:

- among the revenues, FO must include the bid he would offer within the public auction, which is not the case in (5);
- similarly to the CTO's behavior, public auction's rules assure that the expected profit μ_j is a truly revealed value (see Myerson (1981)).

Once he has decided where to locate his facility, the FO can decide on the type of plant to be chosen, therefore determining the fixed costs f_j and the variable cost ct_j. It should be noted, however, that the decision on f_j is influenced by the amount of incentives received by PB. It is also quite interesting to note that generally $\bar{\mu}_j$ is for the FO an interaction variable since it depends on how he decides to locate his facility with respect to his competitors'same type of decisions (see Gabszewicz (1996)). In other words, it is a problem of market share. The amount of recovered materials to be sold to firms p_{jl} and its selling price p_j is determined by market rules through negotiations with the firms themselves. These firms will be willing to buy a recovered material instead of a new one if the price will be competitive. We can assume market firms having a demand function for recovered product and materials such as:

$$q_{jl} = q_{jl}(p_j) \quad \frac{dq}{dp} < 0 \tag{7}$$

where:

$$q_{jl} = \begin{cases} q_{jl} = q_{jl}(p_j) \text{ if } \quad p \in]0, p_M[\\ 0 \text{ otherwise} \end{cases}$$

In the formula p_M is the market price for buying a new product on the market instead of buying a recovered one. If the recovered product costs less then the "new" one there exists a downsloped demand function by market firms. If p_j is bigger than the market price p_M, firm demands are zero since firms prefer to buy new products on the market.

At this point, it is worth to focus our analysis on the coordination mechanisms emerging from the economic behaviors, in particular from the interrelationships among agents - the interaction variables. We aim at identifying the coordination mechanisms determining the specific value of each interaction variable. We present a scheme summarizing the interaction variables and the interrelationships among agents (see figure 4).

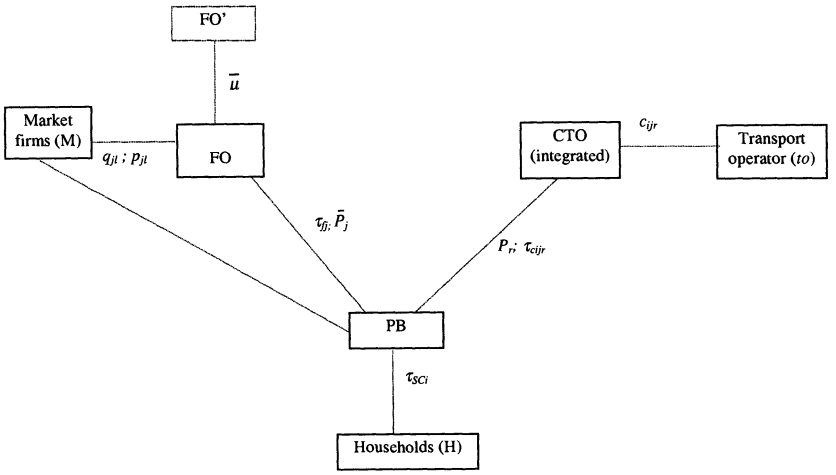


Fig. 4. Main interaction variables of the reverse logistics system

By considering the set of interrelationships among agents and the interaction variables we can identify the main coordination mechanism as in following table 1.

The following coordination mechanisms emerge from table 1:

- $\mathcal{PA}$: principal-agent mechanisms. These regulate the decisions about the incentives by PB which in turn influence the decision of the selling price of recovered materials, the investment costs for the recovery facilities, the collection/transportation costs and the amount of waste produced by households;

Table1. Interaction variables, interdependencies among agents and coordination mechanisms

	$\bar{\mu}_j$	q_{jl}	p_j	τ_{P_j}	P_j	τ_{f_j}	P_r	$\tau_{c_{ijr}}$	τ_{SC_i}	c_{ijr}
M-FO		Market	Market							
FO-FO	Comp. location									
CTO-to										Market
PB-FO						Principal -agent	Auction			
PB-M PB-M				Principal -agent						
PB-CTO							Auction	Principal -agent		
PB-H									Principal -agent	

- $\mathcal{M}$: market mechanisms. They determine the selling price of the recovered materials, the quantity of recovered materials sold on the market and the pick-up/transportation costs;
- $\mathcal{A}$: auction rules mechanisms. They determine th auction bids P and $\bar{P}$ and therefore influence the expected profit of the operators;
- $\mathcal{CL}$: competitive location mechanisms. Such mechanisms are mainly related to game theory and they determine the strategic decision by FO on where locate their plants (and therefore π_{3j} and π'_{3j}).

An illustration of the coordination mechanism is that of table 2.

Table2. Coordination mechanisms and interaction variables

	$\bar{\mu}_j$	q_{jl}	p_j	τ_{P_j}	P_j	τ_{f_j}	P_r	$\tau_{c_{ijr}}$	τ_{SC_i}	c_{ijr}
$\tau(sd)$				•		•		•	•	
p		•	•							•
d					•		•			
fo	•									

5 Model Specification and Conclusions

On the basis of the above mentioned analysis the following "open" specification of the general framework model can be proposed:

$$\pi_1 = \tau_{SC} X_{SC} + \sum_r P_r + \sum_j \bar{P}_j + \tau_{f_j} \sum_j f_j + \tau_{c_{ijr}} \sum_i \sum_j \sum_r x_{ijr} c_{ijr} +$$

$$+ \tau_{p_j} \sum_j \sum_l p_j q_{jl} - TX_R \tag{8a}$$

$$\pi_{2r} = P_r(\mathcal{A}) - FC_r - VC_r\{K_r, I, a_{kr}, c_{ijr}(\tau_{c_{ijr}}, \mathcal{M}), x_{ijr}[x_{SC_i}(\tau_{SC_i})]\}$$
$$= \mu_r(\mathcal{A}) \tag{8b}$$

$$\pi_{3j} = R_j[q_{jl}(\mathcal{M}), p_j(\tau_{p_j}, \mathcal{M}), \bar{P}_j(\mathcal{A})] +$$
$$- f_j(\tau_{f_j}) - VC_j[ct_j, x_{ijr}(\mathcal{CL})] = \bar{\mu}_j(\mathcal{A}) \tag{8c}$$

$$\pi'_{3j} = R_j[q_{jl}(\mathcal{M}), p_j(\tau_{p_j}, \mathcal{M})] +$$
$$- f_j(\tau_{f_j}) - VC_j[ct_j, x_{ijr}(\mathcal{CL})] = \bar{\mu}_j \tag{8d}$$

$$c_{ijr} = \{\mathcal{M}_{CTO,to}\} \tag{8e}$$

$$P_r = \{\mathcal{A}_{PB,CTO}\} \tag{8f}$$

$$\bar{P}_j = \{\mathcal{A}_{PB,FO}\} \tag{8g}$$

$$\tau_{c_{ijr}} = \{\mathcal{PA}_{PB,CTO}\} \tag{8h}$$

$$\tau_{SC_i} = \{\mathcal{PA}_{PB,H}\} \tag{8i}$$

$$\tau_{f_j} = \{\mathcal{PA}_{PB,FO}\} \tag{8j}$$

$$q_{jl} = \{\mathcal{M}_{FO,M}\} \tag{8k}$$

$$p_j = \{\mathcal{M}_{FO,M}\} \tag{8l}$$

$$\tau_{p_j} = \{\mathcal{PA}_{PB,M}\} \tag{8m}$$

$$\bar{\mu}_j = \{\mathcal{CL}_{FO,FO'}\} \tag{8n}$$

$$\Pi = \phi(\pi_1, \pi_2, \pi_3, \pi'_3) \tag{8o}$$

For the model to be completely specified we would need to find analytical representations for the coordination mechanisms. This is a kind of work still lying ahead.

The analysis carried out so far has shown that a final configuration of a logistics system can not be thought as a sequence of problems, even if they are complex, solved by a single decision-maker. Indeed, it should come out from the interactions of the behaviors of different economic agents. This brings us to compare the SAS and MAS approaches. In particular, the question is: what should be the integrations or variations that the behavioral approach induced to the classical one? The bottom line is that due to the identification of a certain number of agents and interrelationships regulated by some coordination mechanisms we identify *a larger spectrum of system problems and activities which need a modelling framework* to be analysed and explained. Considering a reverse logistics system, the comparison between the two approaches can be done as in the following table (see 3).

Further significant results of the analysis seem to be:

- in the reverse logistics system which has been the subject of the case-study, there are a relevant number of interaction variables as

Table3. SAS-MAS comparison in modelling terms

Logistics problems/ activities	SAS approach Classic models	MAS approach Agents involved	MAS approach Behavioral models
Location of incinerators, composting and landfilling plants	UFPL, CFPL	Public bodies	hierarchical multi--criteria models
Location of processing (material recovery) plants	UFPL, CFPL	Facility operators	Economic models of location (*competitive location, price-sensitive*)
Assignment of collection and transportation logistics services		Local authorities-CTO	auction models, principal-agent model
Assignment of processing logistics services		Local authorities-FO	auction models, principal-agent model
Operation of collection and transportation logistics services (1)		CTO-to	market mechanism
Operation of collection and transportation logistics services (2)	VRP, VRSP, VRSPTW, loading, fleet composition mode choice	CTO	Classical optimization models, possibly reformulated
Provision of incentives for the logistics services		Local authorities-households/FO/ /CTO/market firms	principal-agent models

variables characterizing the economic behavior of the agents. This means that the classical interpretation of the decision-making process when developing a Decision Support System, which is based on the identification, on one hand, of the variables that can be completely decided by the agents and, on the other hand, of those that should be taken as given, seems to be oversimplified. On the contrary, we have found out that the bulk of the relevant variables of a logistics system lies within the so-called interaction variables. This stresses the need to deepen the analysis on the coordination mechanisms in order to understand how such variables are determined within logistics systems.

- if we go beyond the classical approach of logistics system analysis as it is seen in terms of a planning process based on optimization problems, we realize the crucial role of the public bodies (and eventually of the State) as regulators of an economic and logistics system. Since they no longer plan and control the system in a centralized manner, they aim at promoting and stimulating, on one hand, the generation of potential reverse flows by system agents (i.e., households) and, on the other hand, the profitability of the activities of the agents of the reverse logistics system (i.e., CTO, FO and market firms). In this way, PB helps the reverse logistics system to work better and be more efficient in a social way. Therefore, from a normative standpoint, instead of optimally and deterministically producing a final configuration of the system, we should strength the analysis on how public authorities should "regulate" a decentralized systems by means of incentives, subsidies, taxes, etc. so as to achieve an optimal or satisfying configuration in terms of social utility . Moreover, in addition to tools such as incentives and taxes, it is quite clear that the choice of a given coordination mechanism regulating some interdependencies among agents is capable of strongly influenced the global performance of the system. Therefore, such a choice can be seen as a further tools in the public authorities'hands in order to achieve some social objectives. Once

again, there is the need to better understand the economic rationale at the core of the coordination mechanisms.
Even if we did not empirically address the issue of the non-urban waste logistics, that is, how a private company organizes its forward-reverse logistics systems, it seems that the above considerations still hold. In other words, given that private companies often do not own and manage "directly" its logistics systems rather it coordinates with other independent agents (suppliers, agencies, importers, exporters, franchisors, etc.) there is a problem for it of how to create and structure incentive schemes for the independent agents in order to achieve some company's goals.

To conclude, the evolution of the logistics and economic system analysis makes it clear that a strong accent should be put on the specification of the different actors involved in the system and their behavioral characteristics (disaggregated approach). From this the need to study the interrelationships and coordination mechanisms among the actors stems directly. From a modelling standpoint, theorists should produce efforts in order to put forward a new generation of models capable of interpreting how the systems work and how an optimal configuration could be reached at the end. As already said, such models should have the characteristics of incorporating both specific behavioral hypothesis (coming out from empirical analyses) and coordination mechanisms. The efforts for producing such models is certainly a future line of research.

References

Ayres, R.U. (1997): Metals recycling: economic and environmental implications. INSEAD Working Paper 97/59/EPS/TM.

Ayres, R.U. (1998): The second law, the fourth law, recycling and limits to growth. INSEAD Working Paper 98/38/EPS/CMER.

Ayres, R.U. / Ferrer, G. / Van Leynseele, T. (1997): Eco-efficiency, asset recovery and remanufacturing, INSEAD Working Paper 97/35/EPS/TM.

Arrow, K. / Debreu, G. (1954): Existence of an equilibrium for a competitive economy, *Econometrica.*

Baligh, H.H. (1986): Decison rules and transactions, organizations and markets. *Management Science*, 32, 1480-1491.

Ballou, R. H. (1999): *Business Logistics Management.* New Jersey, Prentice Hall.

Balsmeier, P.W. / Voisin, W.J. (1996): Supply Chain Management: a time-based strategy. *Industrial Management*, 38, 24-27.

Barros, A.I. / Dekker, R. / Scholten, V. (1998): A two-level network for recycling sand: a case study. *European Journal of Operational Research* 110, 199-214.

Bechtel, C. / Jayaram, J. (1997): Supply Chain Management: a strategic perspective. *The International Journal of Logistics Management* 8(1), 15-34.

Bianco, L. (1987): Mathematical models in logistic system design. In *Freight Transport and Logistics*, Proceedings, Bressanone, July 1987.

Bloemhof-Ruwaard, J.M. / Fleischmann, M. / van Nunen, Jo A.E.E. (1999): Distribution Issues in Reverse Logistics, in *New Trends in Distribution Logistics*, Lecture Notes in Economics and Mathematical Systems 480 (M. Grazia Speranza, Paul Stahly (Eds.)), Springer, 23 - 44

Boulding, K. (1966): The economics of the coming spaceship Earth. in *Environmental quality in a growing economy* Baltimore, Johns Hopkins university Press.

Cooper, M.C. / Douglas, M.L. / Pagh, J.D. (1997): Supply Chain Management: more than a new name for logistics. *The International Journal of Logistics Management* 8(1), 1-14.

Daly, H.E. (1980): *Economy, ecology, ethics: essay toward a steady-state economy*, S.Francisco, W.H. Freeman.

Daly, H.E. (1986): Thermodynamic and economic concepts as related to resource-use policies: comment. *Land Econ.* 62, 319-322.

Davis, R. / Smith, R.G. (1983): Negotiation as a metaphor for distributed problem solving. *Artificial Intelligence*, 20, 63-109.

Ferrer, G. (1996a): Market segmentation and product line design in remanufacturing. INSEAD Working Paper 96/66/TM.

Ferrer, G. (1996b): Parts recovery problem: the value of information in remanufacturing. INSEAD Working Paper 96/63/TM.

Ferrer, G. (1996c): The economics of tire remanufacturing. INSEAD Working Paper 96/39/TM

Ferrer, G. (1997): The economics of PC remanufacturing. INSEAD Working Paper 97/38/TM.

Ferrer, G. / Ayres, R.U. (1998): The impact of remanufacturing in the economy. INSEAD Working Paper 98/14/EPS.

Fleishmann, M. / Bloemhof-Ruwaard, J.M. / Dekker R. / Van der Laan, E. / Van Nunen, J.A.E.E. / Van Wassenhove, L.N. (1997): Quantitative models for reverse logistics: a review. *European Journal of Operational Research* 103, 1-17.

Georgescu-Roegen, N. (1971): *The entropy law and the economic process.* Harvard University Press, Cambridge.

Johnson, F.P. (1998): Managing value in reverse logistics system. *Transportation Research - E (Logistics and Transportation Review)*, vol.34, No. 3, pp.217-227.

Kroon, L. / Vrijens, G. (1995): Returnable containers: an example of reverse logistics. *International Journal of Physical Distribution and Logistics Management* 25(2), 56-68.

Kurose, J.F. / Simha, R. (1989): A microeconomic approach to optimal resource allocation in distributed computer system, *IEEE Transaction on Computers*, 38(5), 332-353.

Lumer, E. / Huberman, B.A. (1990): Dynamics of resource allocation in distributed systems SSL-90-05. Palo Alto, CA, Xerox PARC.

Malone, T.W. (1987): Modeling coordination in organizations and markets. *Management Science*, 33, 1317-1332.

Malone, T.W. (1988): What is coordination theory? Working Paper #2051-88. Cambridge, MA, MIT Sloan School of Management.

Malone, T.W. (1992): Analogies between human organization and artificial intelligence systems: two examples and some reflections. In M. Masuch (ed.)*Distributed Intelligence: Perspectives of Artificial Intelligence on Organization and Management Theory.* Amsterdam, Elsevier.

Malone, T.W. / Crowston, K. (1994): The interisciplinary study of coordination. *ACM Computing Surveys*, 26(1), 87-119.

Malone, T.W. / Crowston, K.G. (1991): Toward an interdisciplinary theory of coordination. Technical Report #120, Cambridge, MA, Massachusetts Institute of Technology, Center for Coordination Science.

Malone, T.W. / Smith, S.A. (1988): Modeling the performance of organizational structures. *Operations Research*, 36(3), 421-436.

Malone, T.W. / Crowston, K. / Lee, J. / Pentland, B. (1993): Tools for inventing organizations: toward a handbook of organizational processes. In Proceeding of the 2nd IEEE Workshop on Enabling Technologies Infrastructures for Collaborative Enterprises. Morgantown, WV, April 20-22.

Miao, X. / Luh, P.B. / Kleinman, D.L. (1992): A normative-descriptive approach to hierarchical team resource allocation. *IEEE Transaction Systems, Man and Cybernetics*, 22(3), 482-497.

Myerson, R.B. (1981): Optimal auction design. *Mathematics of Operations Research*, 6, 58-73.

Motwani, J. / Larson, L. / Ahuja, S. (1998): Managing a global supply chain partnership. *Logistics Information Management* 11(6), 349-354.

Nash, C.A. / Whiteing, A.E. (1987): Mode choice: a total distribution approach. In: *Freight Transport Planning and Logistics*, Springer-Verlag, Bressanone.

Pearce, D.W. / Turner, R.K.(1990):*Economics of natural resources and the environment.* Harvester Wheatsheaf, Hertfordshire.

Pohlen, T.L. / Farris, M. (1992): Reverse logistics in plastic recycling. *International Journal of Physical Distribution and Logistics Management* 22(7), 35-47.

Reiter, S. (1986): Informational incentive and performance in the (new)2 welfare economics. In S. Reiter (eds) Studies in *Mathematical Economics (Studies in Mathematics, Volume 25).* Mathematical Association of America.

Rose, J. (1994): Waste management and life-cycle analysis. *Environmental Management and Health* 5(1), 5-6.

Ross, S. (1973): The economic theory of agency. *American Economic Review*, 63, 134-139.

Schneeweis, C. (1999): *Hierarchies in distributed decision making.* Springer-Verlang.

Sodhi, S. M. / Reimer, B. (2001): Models for recycling electronics end-of-life products. *OR Spektrum* 23(1), 97-115.

Sikora, R. / Shaw, M.J. (1998): A multi-agent framework for the coordination and integration of information systems. *Management Science*, vol. 44, n. 11, 65-78.

Thierry, K. / Salomon, M. / Van Nunen, J. / Van Wassenhove, L. (1995): Strategic issues in product recovery management. *California Management Review* 37(2), 114-135.

Gabszewicz, J.J. / Thisse, J. (1996): Spatial competition and the location of firms. In *Regional and Urban Economics*, edited by Richard Arnott, Harwood Academic Publishers.

Van Brussel, H. / Valckenaers, P. / Bongaerts, L. / Wyns, J. (1997a): Architectural and system design issues in holonic manufacturing systems. Second K.U. Leuven Tutorial on Holonic Manufacturing, Leuven, Belgium, 12 Sept.

Van Brussel, H. / Wyns, J. / Valckenaers, P. / Bongaerts, L. / Peeters, P. (1997b): Reference architecture for holonic manufacturing systems. Second K.U. Leuven Tutorial on Holonic Manufacturing, Leuven, Belgium, 12 Sept.

Van der Laan, E.A. / Fleishmann, M. / Dekker, R. / Van Wassenhove, L.N. (1998): Inventory control for joint manufacturing and remanufacturing. INSEAD Working Paper 98/61/TM.

Van Parunak, H.V.D. / Ward, A. / Fleischer, M. / Sauter, J. (1997a): A marketplace of design agents for distributed concurrent set-based design. ISPE/97: Fourth ISPE International Conference on Concurrent Engineering: Research and Applications, Troy, Michigan, August 20-22.

Van Parunak, H.V.D. / Sauter, J. / Clarke, S. (1997b): Toward the specification and design of industrial synthetic ecosystem. Fourth International Workshop on Agent Theories, Architectures and Language (ATAL '97).

Williamson, O.E. (1975) *Markets and Hierarchies.* New York, Free Press.

Performance of MRP in Product Recovery Systems with Demand, Return and Leadtime Uncertainties

Christian Gotzel and Karl Inderfurth

Faculty of Economics and Management, Otto-von-Guericke University of Magdeburg, Germany

Abstract. In this paper the performance of an extended MRP approach (MRRP) for a hybrid single-stage production/remanufacturing system with external return flows is examined. All costs are assumed to be strictly proportional so that a lot-for-lot ordering policy applies. A scenario with stochastic demands and returns and deterministic processing times of equal length for both production and remanufacturing forms the basis of the analysis. Additionally, cases of unequal and stochastic processing times are considered. It turns out that application of MRRP leads to near-optimal results. MRRP even outperforms optimized pull control rules from Stochastic Inventory Control (SIC) if the leadtimes differ considerably.

1 Introduction

In the last decade due to both economic incentives and legal pressure firms more and more are going to organize product and material cycles, by this way supporting the approach of an "economy of circulation". As a result, product recovery issues become more and more important. The term product recovery refers to activities for regaining materials and value added of used products. This especially holds for the recovery option of remanufacturing by which used products are brought up to an 'as new' condition so that the quality standards of new products are fulfilled (see Thierry et al. (1995)).

Since traditional production planning was not intended to handle reverse material flows, new challenges arise for the development of new planning instruments. The complexity of product recovery management arises from three problems. At first, integration of production and remanufacturing requires guidelines for the use of both alternative supply modes. Second, a cost trade-off between remanufacturing and disposal that depends on the stochastic return flow has to be taken into account. And third, product recovery is accompanied by additional uncertainties with respect to quantity, quality and arrival time of the returned goods. In this context, also significant uncertainty concerning the processing time for remanufacturing can be observed. Therefore, planning approaches should take these specific uncertainties into consideration and provide measures to cope with them (see

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

Guide (1998)). In this paper we address the question of how to control production, remanufacturing, disposal and inventories under stochastic conditions in a quite simple, but near-optimal way.

In the literature we find different approaches to tackle this material coordination problem. Fleischmann et al. (1997) give an overview of quantitative models for reverse logistics. Van der Laan et al. (1999) examine SIC policies for a product recovery problem. An MRP based planning approach for the problem under consideration is given in Inderfurth and Jensen (1999). The latter approach has the advantage to be easily applicable for planning problems with general and complex product structures. However, in order to be recommendable it should at least work successfully in a single-level environment. This paper will clarify that this condition indeed is fulfilled.

2 Model Assumptions

In the following we consider a single-stage production/remanufacturing problem. In each period of a specified planning horizon requirements are fulfilled from a serviceable products (SP) inventory which can either be filled from regular production or from remanufacturing of returned products. These returns arrive at the end of a period. At the beginning of the next period one has to decide how many new items to produce, how many returned items to remanufacture and how many to dispose of. Additionally, returns can be temporarily stocked in the remanufacturables (RP) inventory. For both production and remanufacturing specific lead times have to be considered. All costs are assumed to be strictly proportional. Unsatisfied demand is backordered. Both demand and returns in each period are assumed to be stochastic. The same holds for the remanufacturing leadtime.

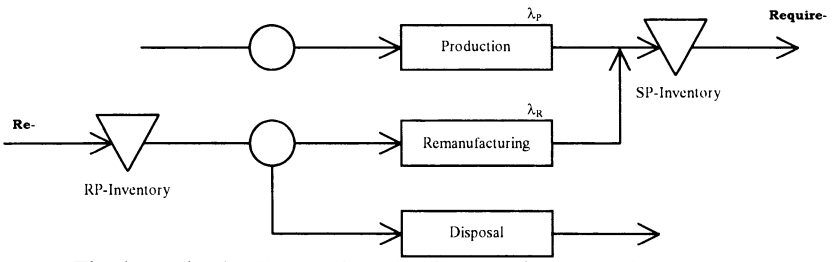


Fig. 1: Production/Remanufacturing System with External Return Flows

The notation used can be summarized as follows:

c_P production cost per unit
c_R remanufacturing cost per unit
c_D disposal cost per unit
h_S holding cost for serviceables per unit and period

h_R holding cost for remanufacturables per unit and period
ν shortage cost for serviceables per unit and period
λ_P production leadtime in periods
$\tilde{\lambda}_R$ remanufacturing leadtime in periods (stochastic)
$\tilde{GR}_t$ gross requirements in period t (stochastic)
$\tilde{R}_t$ returns in period t (stochastic)

The respective expectations and standard deviations of a stochastic variable $\tilde{z}$ are denoted by μ_z and σ_z.

POP_t planned order release production at (the beginning of period) t
POR_t planned order release remanufacturing at t
POD_t planned order release disposal at t
SOH_t serviceables on hand at the end of period t
SST safety stock

A periodic review control system is employed. We assume that MRP is applied in a rolling schedule framework where replanning takes place after each period. Control procedures are searched that keep expected total cost per period as small as possible.

3 Control Approaches

Under the preliminary assumption of identical and deterministic leadtimes, i.e. $\lambda = \lambda_P = \lambda_R$, the optimal control policy for each period is a simple SIC-type pull policy with three control parameters (see Inderfurth (1997)): a produce-up-to level S, a remanufacture-up-to level M and a dispose-down-to level D. This so-called (S,M,D)-policy has the following structure:

$$POP_t = \max\{S - x_{Et}, 0\} \tag{1}$$

$$POR_t = \min\{x_{Rt}, \max\{M - x_{St}, 0\}\} \tag{2}$$

$$POD_t = \max\{x_{Rt} - \max\{D - x_{St}, 0\}, 0\} \tag{3}$$

The decisions are based on comparisons of the fixed control parameters with different inventory positions which are defined as follows:

x_{Rt} remanufacturables on hand at the beginning of period t

$x_{St} = SOH_{t-1} + \sum_{i=1}^{\lambda} POP_{t-i} + \sum_{i=1}^{\lambda} POR_{t-i}$ inventory position of serviceable products at the beginning of period t

$x_{Et} = x_{St} + x_{Rt}$ echelon inventory position at the beginning of period t

Although we know the structure of the optimal policy we cannot give closed-form expressions for the optimal control parameters. To calculate these parameters, one has to rely on numerical methods.

An alternative approach for planning is to extend traditional MRP control to a material requirements and recovery planning (MRRP). In MRP all inputs are considered as quasi-deterministic, e.g. estimated by their expected values. In a first step, the planning procedure uses inventory information to derive net requirements and planned order releases from a gross requirements schedule, and in the next step orders are time-phased according to the given leadtime information. To cover uncertainties, safety buffers in form of safety stock or safety leadtimes can be applied. Furthermore, a rolling horizon planning allows to react to unexpected changes in demand and supply.

For extending MRP for the product recovery situation some adjustments become necessary. At first, we need a priority rule for the net requirements fulfillment. Since a returned item which is not remanufactured has to be disposed of at cost c_D, remanufacturing is profitable if $c_R \leq c_P + c_D$. In this case, remanufacturing is preferable to regular production and therefore the priority rule is to first use all returns available for the net requirements fulfillment before using production. Second, a decision rule for disposal of returns is needed. If there is such a large surplus of returned items that the respective holding costs exceed the remanufacturing cost benefit $c_P + c_D - c_R$ then disposal becomes profitable. Therefore, the critical runout time for excess returns is given by

$$\tau = \left\lfloor \frac{c_P + c_D - c_R}{\min\{h_R, h_S\}} \right\rfloor \tag{4}$$

which we can use to determine a stock level of returns (*DST*) that must not be exceeded:

$$DST = \max\left\{ \max_{1 \leq j \leq \tau} \left\{ \sum_{i=1}^{j} \mu_{GR,t+\lambda+i} - \sum_{i=1}^{j} \mu_{R,t+i-1} \right\}, 0 \right\} \tag{5}$$

In the situation of level demands and returns this expression simplifies to:

$$DST = \tau \cdot (\mu_{GR} - \mu_R). \tag{6}$$

Another parameter of MRRP control is the safety stock *SST* which is set up to cover demand uncertainties during the replenishment cycle expressed by the standard deviation of the leadtime demand

$$\sigma_{GR,\lambda} = \sqrt{(\lambda + 1)\sigma_{GR}^2}\,. \tag{7}$$

For minimizing the expected holding and shortage costs the safety stock is now determined using the expression known from the newsboy problem where the

safety factor k depends on the demand distribution (which here is assumed to be a standard normal one with standardized distribution function $\Phi_{N(0,1)}$) and the holding and shortage costs:

$$SST = k \cdot \sigma_{GR,\lambda} \quad \text{with} \quad k = \Phi^{-1}_{N(0,1)}\left(\frac{v}{h_S + v}\right) \tag{8}$$

In our numerical examination we consider three parameter settings given in Table 1 for the case of equal and deterministic leadtimes. Setting (I) includes an equal safety stock for the production and remanufacturing decision and a disposal stock calculated in exactly the way as described above. When returns are stochastic then it can happen that the net returns in a period $\tilde{Z}_t = \tilde{R}_t - \tilde{GR}_t$ take on positive values what is not taken into account in the production ordering decision and results in increased total inventories. To take care of this problem, we introduce an adjustment of the safety stock level SST_P for the production decision in setting (II). This adjustment ΔS equals the expectation of excess inventory. Returns which exceed demands during a certain time span can be used to replace production of future periods and thereby increase the mean level of demand that is satisfied by remanufactured products. Concerning the disposal decision this means that the impact of excess returns over τ periods has to be taken into consideration in an appropriate way.

Table 1: MRRP parameter settings

Setting (I)	$SST_P^I = k_P \cdot \sigma_{GR,\lambda}$ $SST_R^I = SST_P^I$ $DST = \tau \cdot (\mu_{GR} - \mu_R)$	
Setting (II)	$SST_P^{II} = SST_P^I - \Delta S$ $SST_R^{II} = SST_R^I$ $DST^{II} = DST^I + \Delta D$	$\Delta S = \int_0^\infty Z \cdot \varphi_Z^1(Z) \cdot dZ$ $\Delta D = \int_0^\infty Z \cdot \varphi_Z^\tau(Z) \cdot dZ$
Setting (III)	$SST_P^{III} = SST_P^{II}$ $SST_R^{III} = k_R \cdot \sigma_{GR,\lambda}$ $DST^{III} = DST^{II}$	 $k_R = \Phi^{-1}_{N(0,1)}\left(\frac{v}{h_S - h_R + v}\right)$

Using the disposal level given in expression (6) yields an overestimation of necessary disposal orders if positive cumulated net returns are realized during the critical runout time. Thus an adjustment ΔD of the disposal stock level is used to increase DST by the expected positive net returns over τ periods. In setting (III) we

apply a distinct safety stock SST_R based on a remanufacturing-specific safety factor k_R which we can derive from a modified newsboy formulation where we take into account that remanufacturing of returns only generates an increase in holding costs of $h_S - h_R$ per unit. In Table 1 $\varphi_Z^n(Z)$ stands for the density function of net returns Z cumulated over n periods. As a basic result from an analysis of the MRRP control rule described above, Inderfurth and Jensen (1999) show that applying this rule leads to a special case of the *(S,M,D)* policy structure in (1) to (3) which turned out to be optimal in the SIC context. However, the MRRP parameter settings will not necessarily be identical with the optimal SIC parameter values.

4 MRRP Performance for Deterministic Leadtimes

In this section we consider a situation where the leadtimes for production and remanufacturing are equal and deterministic. Requirements and returns are assumed to be normally distributed. A stationary problem with an infinite planning horizon is considered.

At first, we will introduce a reasonable base case scenario. The basic settings of cost, leadtime, and distribution parameters are the following:

$c_P = 20$	$h_S = 3$	$\lambda_P = 10$	$\mu_{GR} = 10$	$\sigma_{GR} = 2$
$c_R = 16$	$h_R = 2$	$\lambda_R = 10$	$\mu_R = 8$	$\sigma_R = 2$
$c_D = 3$	$\nu = 27$			

By systematic variation of parameters we investigate their impact on the MRRP performance. From the cost parameters we see that remanufacturing is profitable and that the critical runout time is τ=3 periods. The returns fraction, i.e. the expected fraction of requirements that is returned, is set at 80%. Holding and shortage costs are fixed in such a way that the optimal non-stockout probability (or service level) reaches 90%. In our investigation optimal SIC parameters are found by a search procedure using stochastic simulation for calculating expected costs. A direct numerical approach for parameter optimization is found in Kiesmüller and Scherer (2000). The respective results for the base case are given in Table 2.

Table 2: Base case results

	S	M	D	TC	RC	SRC
SIC	117.921	122.088	124.683	203.40	35.40	100.00
MRRP (I)	118.501	118.501	124.501	204.21	36.21	102.28
MRRP (II)	118.102	118.501	124.762	204.00	36.00	101.68
MRRP (III)	118.102	121.958	124.762	203.42	35.42	100.04

Since in the identical leadtime case both, the MRP and SIC policy, are equivalent we present the SIC policy parameters to characterize the optimal policy and the MRRP solutions. To make a meaningful comparison between optimal and MRRP costs, we have to suppress the impact of unavoidable cost components. Thus we will only consider the fraction of total costs (TC) that can be influenced by the de-

cisions. To get these relevant costs (RC) the unavoidable fixed costs for production and remanufacturing are subtracted from TC: $RC = TC - (c_P \cdot (\mu_{GR} - \mu_R) + c_R \cdot \mu_R)$. The last column of Table 2 shows the standardized relevant costs (SRC) as our performance measure which represents the RC calculated relatively to the minimal costs from using the optimal SIC rule.
From the results given in this table it can be seen that the maximal cost deviation in the base case stems from MRRP (I) and amounts to 2.3%. For MRRP(III) the policy parameters are very close to the optimal ones resulting in a very small cost deviation. It should be noted that the cost improvement achieved with MRRP(II) and MRRP(III) mainly results from the adjustment of safety stocks SST_P and SST_R whereas the impact of ΔD is rather small.

In the following, the results of a numerical investigation with a systematic variation of a number of model parameters is presented to show the impact of these factors on the optimal SIC parameters and on the MRRP cost performance. The first factor to be considered is the relative remanufacturing cost $\rho = c_R/(c_P + c_D)$. In the base case we have $\rho \tilde{} 0.7$.

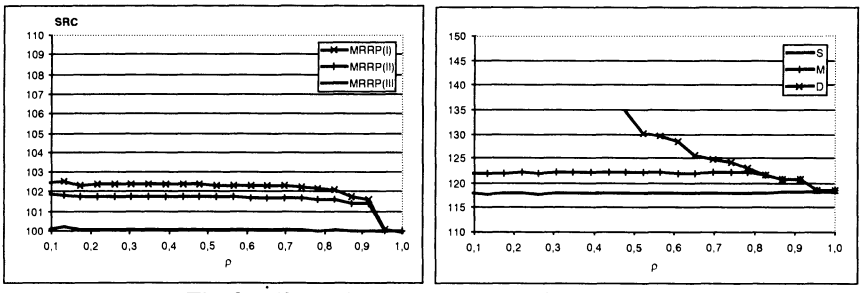


Fig. 2: Effects of the relative remanufacturing cost

Fig. 2 shows the impact of the relative remanufacturing cost. In the left-hand diagram it is depicted that the cost performance is hardly affected by changes of ρ and the MRRP performance increases in the order of the three parameter settings. Only if remanufacturing becomes uneconomic performance switches to optimality with each setting. The right-hand diagram clarifies the impact of ρ on the optimal SIC parameters S, M and D. If remanufacturing costs are small then the disposal level increases and more excess returns are kept in the inventories. On the other hand we end up with a single parameter S-policy if remanufacturing becomes uneconomic i.e., ρ tends to 1. For small parameter values of ρ (here ρ<0.5) the optimal dispose-down-to-level D tends to infinity and has almost no impact on the TC value. For that reason a precise numerical calculation via the simulation approach with its extreme time consumption has not been carried out. So the respective part of the D-curve in Fig. 2 (and analogously in other figures) is omitted.
The holding cost rate $\eta = h_R / h_S$ is an important factor which affects the amount of value added in the remanufacturing process. If the holding cost for remanufacturables is high, i.e. much capital is tied up in the returned goods, then stock-

keeping of remanufacturables becomes uneconomic and optimal control changes to a policy of immediately remanufacturing returns and disposing excess items.

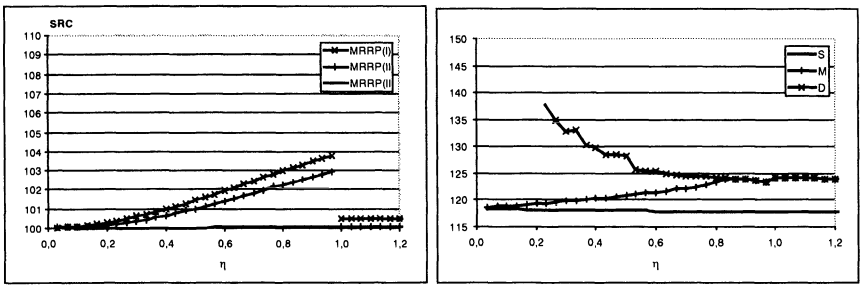


Fig. 3: Effects of the holding cost ratio

This can be seen from the optimal SIC parameters in Fig. 3. For small values of η we have a large RP-inventory but as η increases, more and more stock is shifted from the returns' to the serviceables' inventory. Because this impact has not been incorporated in the MRRP parameter settings (I) and (II), we can notice a deterioration of the cost performance for $\eta \to 1$ whereas MRRP(III) shows no significant deviation from optimum. The holding cost ratio shows a significant interaction with two other factors. To clarify the underlying dependencies we first consider the "service level" $\alpha = v/(h_S + v)$ as the relation of shortage and holding costs.

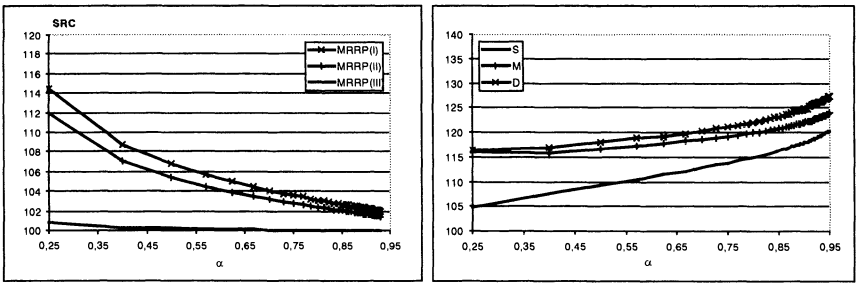


Fig. 4: Effects of the service level

From Fig. 4 we see that the optimal order-up-to level S (or in other words, the MRRP safetey stock) increases with α which is due to the fact that increasing backorder costs are saved by providing additional SP-inventory. On the other hand, optimal control takes advantage of the holding cost difference so that we can notice increasing RP-inventory too. This dependence is taken into account in the determination of the safety factor k_R introduced with MRRP setting (III), resulting in a cost deviation of less than 1% over the whole range of α. The other settings generate too much RP-stock and therefore perform considerably worse, esp. for small α.

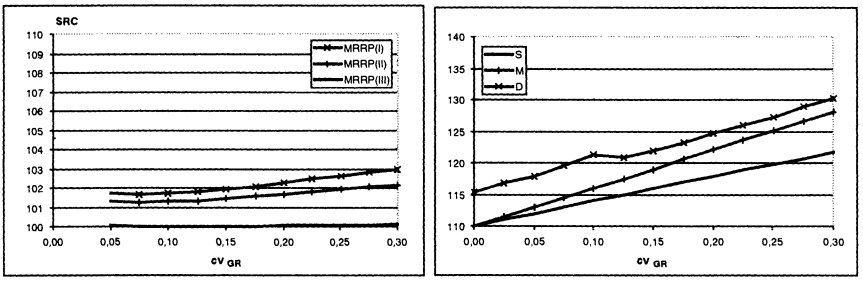


Fig. 5: Effects of the demand variability

Demand variability is another factor that interacts with η. The amount of demand variability, here measured by the coefficient of variation $cv_{GR} = \sigma_{GR} / \mu_{GR}$, has a significant impact on the optimal distribution of returned products between both inventories. With deterministic demand it is profitable to stock excess returns in the RP-inventory. However, if we face highly variable requirements then some safety stock in the SP-inventory is needed to protect against this uncertainty and the RP-holding cost benefit is (at least partially) compensated by backordering costs. Therefore, we can observe a continuous decrease of the optimal RP-inventory when demand uncertainty increases. In Fig. 5, this effect can be seen as the change of the remanufacture-up-to level M from S towards D. The cost comparison reveals that demand variability has no serious effect on costs. As expected, MRRP(III) offers close-to-optimal performance superior to the other settings.

A source of uncertainty that is characteristic for remanufacturing problems is the uncertainty with respect to returns quantities. As a measure of the returns variability we use the coefficient of variation $cv_R = \sigma_R / \mu_R$.

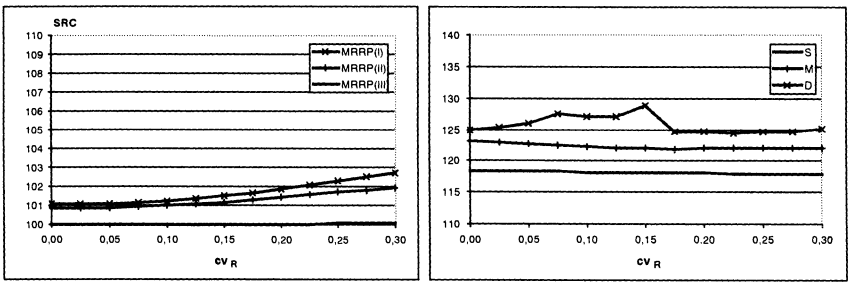


Fig. 6: Effects of the returns variability

However, it is evident that costs are only little affected by this variability. An increase of cv_R results in a higher optimal stock level in both, SP and RP inventories.

A factor that has been shown to seriously affect the MRRP performance (see Inderfurth (1998)) is the returns fraction defined as $\xi = \mu_R / \mu_{GR}$. We see from Fig. 7 that for a wide range of ξ up to 60% all MRRP settings lead to close-optimal costs. For ξ=80%, which represents the base case, all MRRP costs deviate by less

than 2.5% from optimum. For very high return fractions (up to 100%) we find a noticeable performance deterioration of MRRP(I) with costs 10% above the optimum. In contrast, MRRP(III) is still a very good approximation, even for very high returns fractions, and we obtain a maximum deviation from cost optimum of less than 0.4%. From the graphical representation of the optimal SIC parameters we see that for ξ=75% the disposal stock level is reduced to avoid excessive holding costs. If the returns fraction is low then we also find only a small amount of RP-stock due to the restricted availability of returned products.

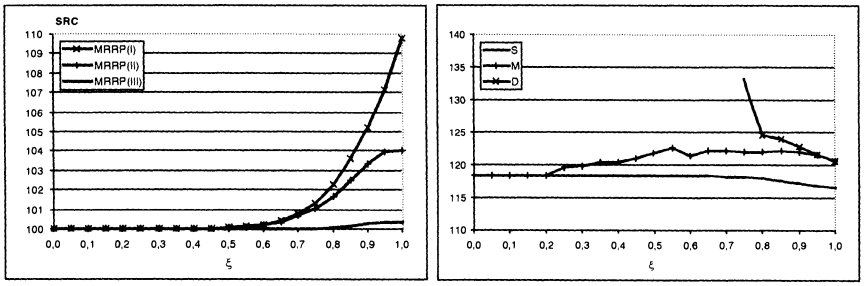


Fig. 7: Effects of the returns fraction

For all parameter constellations it turns out that the parameter setting of MRRP (III) always leads to the best MRRP results and guarantees a cost performance of less than 1% deviation from optimum. Thus, for the remaining analysis we will concentrate on this MRRP variant.

5 MRRP Performance for Unequal Leadtimes

In this section we relax the assumption of identical leadtimes. Under these more general leadtime conditions the decision problem becomes more complicated and creates additional problems for an analysis. At first, it should be noted that we have no precise information about the optimal SIC policy structure, but we know that it will be highly complex (see Inderfurth (1996)). A simple two-parameter (S,D)-policy is optimal if the production leadtime exceeds the remanufacturing leadtime by a single period (see Inderfurth (1997)) but this is restricted to a scenario without stockkeeping of returns.

The unequal leadtimes case is addressed by Inderfurth and Van der Laan (2001) who show that the performance of suboptimal SIC policies can be improved by optimizing the stock information aggregated to the serviceables inventory position. With unequal leadtimes this aggregation can be regarded as a heuristic procedure. We may obtain a better cost performance when using an aggregation of x_{St} which is different from that given in section 3 due to replacing the physical remanufacturing leadtime λ_R by an artificial one denoted by l_R:

$$x'_{St} = SOH_{t-1} + \sum_{i=1}^{\lambda_P} POP_{t-i} + \sum_{i=1}^{l_R} POR_{t-i} \,. \tag{9}$$

Now, the number of periods l_R for which remanufacturing orders are included in the inventory position, is considered as a policy variable called effective remanufacturing leadtime. l_R (with $l_R \leq \lambda_R$) describes how many past remanufacturing orders are effectively considered for the serviceables inventory position.

Under the assumptions of our base case scenario, but considering the described leadtime manipulation, example results from a computational investigation are given in Table 3 in order to show the impact of the physical remanufacturing leadtime on the optimal effective leadtime l_R^*. Hereby λ_R has been varied over the range from 0 to 20 periods whereas the production leadtime is fixed to $\lambda_P = 10$ periods.

Table 3: Search for l_R^*

	l_R																				
λ_R	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	**43.47**																				
1	46.05	**42.86**																			
2	46.68	45.26	**42.25**																		
3	47.42	45.77	44.43	**41.65**																	
4	48.01	46.51	44.93	43.61	**41.12**																
5	48.33	46.99	45.51	44.00	42.71	**40.56**															
6	48.52	47.25	45.93	44.53	43.05	41.70	**39.91**														
7	48.53	47.38	46.13	44.82	43.40	41.96	40.64	**39.08**													
8	48.60	47.43	46.23	45.00	43.66	42.25	40.79	39.45	**37.97**												
9	48.70	47.46	46.28	45.07	43.80	42.46	41.04	39.57	38.11	**36.74**											
10	49.05	47.56	46.28	45.08	43.86	42.54	41.19	39.70	38.24	36.81	**35.40**										
11	50.41	49.07	47.78	46.50	45.29	44.04	42.76	41.42	39.97	38.56	37.18	**36.10**									
12	50.85	49.40	48.24	47.06	45.77	44.56	43.26	41.97	40.64	39.19	37.76	**36.41**	37.11								
13	51.19	49.56	48.46	47.24	46.05	44.75	43.48	42.17	40.87	39.50	37.99	**36.53**	36.92	38.24							
14	51.43	49.70	48.47	47.37	46.11	44.92	43.61	42.28	40.94	39.58	38.16	**36.60**	36.90	37.97	39.44						
15	51.69	49.72	48.50	47.31	46.19	44.92	43.68	42.32	40.99	39.58	38.19	**36.70**	36.91	37.79	39.13	40.61					
16	52.34	49.85	48.46	47.35	46.10	44.96	43.68	42.36	40.98	39.63	38.15	**36.72**	36.98	37.72	38.92	40.29	41.78				
17	57.30	50.17	48.52	47.33	46.19	44.93	43.74	42.40	41.05	39.66	38.24	**36.71**	37.04	37.86	38.85	40.16	41.55	43.31			
18	58.07	50.35	51.62	47.36	46.19	44.99	43.70	42.49	41.10	39.69	38.26	**36.76**	37.02	37.87	38.90	40.06	41.38	42.75	45.60		
19	58.48	55.14	48.63	47.34	46.16	44.96	43.72	42.37	41.14	39.71	38.23	**36.75**	37.05	37.83	38.89	40.15	41.40	42.58	44.02	45.47	
20	58.56	56.63	52.81	47.30	46.13	44.96	43.73	42.43	41.08	39.77	38.31	**36.77**	37.10	37.89	38.87	43.74	46.16	45.97	43.76	45.45	46.48

From the relevant costs given in the Table 3 we see that if the remanufacturing leadtime exceeds the production leadtime by more than one period, an optimized inventory position is based on an effective leadtime of $l_R^* = 11$ periods.

However, apart from the remanufacturing leadtime, the returns fraction also has an influence on l_R^*. This impact is shown in Table 4 where the base case is considered for a leadtime combination of $\lambda_P = 10$ and $\lambda_R = 15$.

Table 4: Search for l_R^*

	l_R					
ξ	10	11	12	13	14	15
0.0	34.68	34.68	34.68	34.68	34.68	34.68
0.1	34.73	**34.70**	34.72	34.74	34.76	34.78
0.2	34.89	**34.77**	34.86	34.95	35.03	35.13
0.3	35.13	**34.89**	35.10	35.31	35.50	35.72
0.4	35.47	**35.06**	35.43	35.80	36.15	36.54
0.5	35.89	**35.27**	35.80	36.38	36.93	37.55
0.6	36.42	**35.54**	36.15	36.94	37.77	38.71
0.7	37.09	**35.94**	36.46	37.40	38.56	39.90
0.8	38.19	**36.70**	36.91	37.79	39.13	40.61
0.9	42.74	39.95	**39.20**	39.41	40.25	41.31
1.0	64.44	53.57	49.50	47.79	45.07	**45.06**

Extending the MRRP approach to the general leadtime case involves some changes in the underlying MRP logic which becomes slightly more complex, but still follows simple MRP computation rules. On the other hand, we need to review the MRRP control parameters.

If the production leadtime exceeds the remanufacturing leadtime then we face additional returns uncertainty for $\lambda_P - \lambda_R$ periods so that the variance of the net reuirements is increased to $\sigma^2_{NR,\lambda} = Var\left\{\sum_{i=0}^{\lambda_P} \tilde{GR}_{t+i} - \sum_{i=1}^{\lambda_P-\lambda_R} \tilde{R}_{t+i-1}\right\}$. Assuming that requirements and returns are i.i.d. then we can express the standard deviation of the net requirements by $\sigma_{NR,\lambda} = \sqrt{(\lambda_P+1)\sigma^2_{GR} + (\lambda_P - \lambda_R)\sigma^2_R}$. This expression will be used to replace the respective formula for equal leadtimes in (7) so that the additional uncertainty is covered by the safety stock implemented in the MRRP parameters. Additionally, in order to reflect the changed leadtime situation a modified critical runout time τ' is now given by: $\tau' = \max\{\tau - \lambda_P + \lambda_R, 0\}$. This way, the leadtime gap affects the determination of *DST* and its adjustment introduced with MRRP(II).

Now that we have introduced extensions for both control approaches the adjusted, but still suboptimal, (*S,M,D*)-policy with optimized parameters is compared to MRRP(III). Fig. 8 shows the cost comparison for a production leadtime of $\lambda_P = 10$ periods and λ_R varied from 0 to 20 periods.

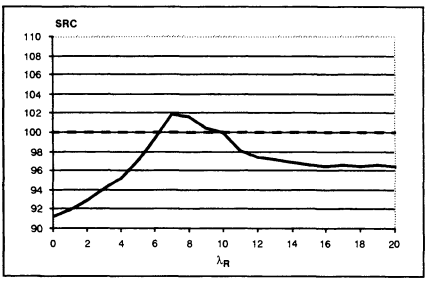


Fig. 8: MRRP performance for uneqal leadtimes

From the diagram, it can be seen that MRRP outperforms the SIC policy if the remanufacturing leadtime exceeds the production leadtime. For $7 \le \lambda_R \le 10$ SIC

results in a better performance compared to MRRP which leads to costs less than 2% above optimum. However, if the remanufacturing leadtime is relatively short ($\lambda_R < 7$) we observe a considerably deteriorating SIC performance. This interesting result can be interpreted as follows: Facing substantial leadtime differences, the way how MRRP approach keeps track of the single order information in a deterministic approach generates less suboptimality than the way SIC incorporates stochastics under neglecting detailed order informations by using aggregate inventory position data.

6 MRRP Performance for Stochastic Leadtimes

An additional source of uncertainty arising from the return of used products is that their quality is not known in advance, so that we have lack of information on how many time is needed for operations like dismantling, quality-checking and processing. In consequence, remanufacturing processing times may be highly stochastic.

It must be noted that in the presence of remanufacturing leadtime variability we have no knowledge of the optimal policy. The simple SIC policy applied is a suboptimal one, but lacking a better alternative we will use it as a benchmark again.

The remanufacturing leadtime $\tilde{\lambda}_R$ is assumed to be a binomially distributed stochastic variable. Furthermore, we assume that remanufacturing orders do not cross. Fig. 9 shows the impact of remanufacturing leadtime variability on the optimized SIC parameters for the base case scenario.

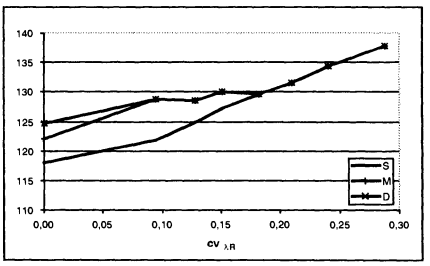


Fig. 9: SIC policy parameters for stochastic leadtimes

As the coefficient of variation cv_{λ_R} increases, additional SP-stock is needed to cover the leadtime uncertainty and to protect against backorder costs, which can be seen from the increasing produce-up-to level S. Therefore, remanufacturing becomes less profitable and stockkeeping of excess returns is reduced which is represented by the remanufacture-up-to level M changing from S towards D. With a higher level of uncertainty we obtain a single parameter S-policy where all excess returns are disposed of.

In order to protect against the leadtime and demand uncertainties we need to set up safety buffers. In MRP systems, application of safety stocks and safety leadtimes is common practice. For evaluating the MRRP performance, we consider the following buffering techniques:

(a) Application of a safety stock to cover demand and leadtime uncertainty. Based on setting (III), we take into account the increased variability of the leadtime demand $\sigma_{GR,\lambda} = \sqrt{(\lambda_P + 1)\sigma_{GR}^2 + \mu_R^2 \sigma_{\lambda_R}^2}$ to determine the safety stock levels SST_P and SST_R.

(b) Application of a safety leadtime to the remanufacturing decision to cover demand and leadtime uncertainty. The safety leadtime is determined using the expression: $l_R^S = \lfloor SST_P / \mu_R + 0{,}5 \rfloor$.

(c) Application of a safety stock (as given in setting (III)) to cover only the demand uncertainty. In combination, a safety leadtime is used to protect just against the leadtime variability. Determination of this safety leadtime $l_R^S = \lfloor (SST_P - SST_p^{III}) / \mu_R + 0{,}5 \rfloor$ is based on the additional SP-stock needed to cover the timing uncertainty.

The left-hand diagram of Fig. 10 illustrates that safety stock buffering (a) leads to costs close to the optimized SIC solution, and is preferable to the application of a safety leadtime that may result in significant deviations from SIC costs.

In the right-hand diagram the cost performance is shown for the case of deterministic demand where only returns and remanufacturing leadtimes are stochastic. Whereas safety stock buffering again leads to a performance close to SIC costs, this is outperformed by the safety leadtime buffering which results in cost savings of up to 17% compared to SIC.

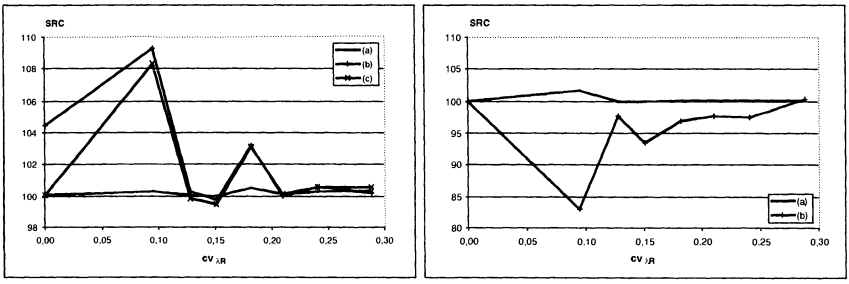


Fig. 10: Effects of remanufacturing leadtime uncertainty

Thus the suggestion is to rely on a safety leadtime as a buffer against the remanufacturing leadtime uncertainty if demand is predictable whereas the safety stock buffer is preferable in the case of demand uncertainty.

7 Summary

The results of our investigation show that application of an MRP-based approach to the production/remanufacturing problem is promising, even in case of multiple stochastic influences. In the first case of equal and deterministic processing times it turns out that an appropriate setting of MRRP control parameters ensures costs very close to optimum. Analyzing the second case of unequal leadtimes reveals that MRRP even outperforms the best known SIC policy if both leadtimes differ considerably. In cases where SIC leads to better results compared to MRRP then we found a maximum cost deviation from optimum of less than 2%. In the third case of stochastic remanufacturing leadtime, MRRP costs are very close to the SIC result if a safety stock is used as a buffer against the timing uncertainty. However, if demand is predictable, then a safety leadtime is suggested to cover the timing uncertainty. Future research will be directed to an evaluation of MRRP performance in the case internal return flows (in form of production rejects or by-products), and to an application of MRRP to a multi-stage case.

Acknowledgement

The research presented in this paper is part of the research on re-use in the context of the EU supported TMR project REVerse LOGistics (ERB 4061 PL 97-0650) in which apart from Otto-von-Guericke University Magdeburg and Eindhoven University of Technology (NL), Aristoteles University of Thessaloniki (GR), Erasmus University Rotterdam (NL), INSEAD (F) and University of Piraeus (GR) take part.
Additionally, this work is supported by a research grant of the Land Saxony-Anhalt.

References

Fleischmann, M./Bloemhof-Ruwaard, J.M./Dekker, R./Van der Laan, E./Van Nunen, J.A.E.E./Van Wassenhove, L.N. (1997): Quantitative Models for Reverse Logistics: A Review. In: European Journal of Operational Research, 103: 1-17.

Guide, V.D.R.,Jr. (2000): Production Planning and Control for Remanufacturing: Industry Practice and Research Needs. In: Journal of Operations Management, 18: 467-483.

Inderfurth, K. (1996): Modeling Periodic Review Control for a Stochastic Product Recovery Problem with Remanufacturing and Procurement Leadtimes. Preprint 2/1996, Faculty of Economics and Management, University of Magdeburg, Germany.

Inderfurth, K. (1997): Simple Optimal Replenishment and Disposal Policies for a Product Recovery System with Leadtimes. In: OR Spektrum, 19: 111-122.

Inderfurth, K. (1998): The Performance of Simple MRP Driven Policies for Stochastic Manufacturing/Remanufacturing Problems. Preprint 32/1998, Faculty of Economics, University of Magdeburg, Germany.

Inderfurth, K./Jensen, T. (1999): Analysis of MRP Policies with Recovery Options. In: Modelling and Decision in Economics, eds. Leopold-Wildburger et al., Physica, Heidelberg-New York, pp. 189-228.

Inderfurth, K./Van der Laan, E. (2001): Leadtimes Effects and Policy Improvement for Stochastic Inventory Control with Remanufacturing. Forthcoming in: International Journal of Production Economics.

Kiesmüller, G.P./Scherer, C.W. (2000): Approximate Optimal Policies for a Stochastic Finite Horizon One Product Recovery Inventory Model. In: Operations Research Proceedings 2000, eds. Fleischmann et al., Springer, Berlin, pp. 310-315.

Thierry, M.C./Salomon, M./Van Nunen, J.A.E.E./Van Wassenhove, L.N. (1995): Strategic Production and Operations Management Issues in Product Recovery Management. In: California Management Review, 37: 114-135.

Van der Laan, E./Salomon, M./Dekker, R./Van Wassenhove, L.N. (1999): Inventory Control in Hybrid Systems with Remanufacturing. In: Management Science, 45: 733-747.

One and Two Way Packaging in the Dairy Sector

Jacqueline M. Bloemhof-Ruwaard[1], Jo van Nunen[1,2], Jurriaan Vroom[3], Ad van der Linden[3] and Annemarie Kraal[3]

[1] Faculty of Business Administration, Erasmus University Rotterdam, 3000 DR Rotterdam, the Netherlands
[2] Deloitte & Touche Bakkenist, 1100 DP Amsterdam, the Netherlands
[3] Logistics Center of Expertise, Campina Melkunie B.V., Woerden, The Netherlands.

Abstract. Choosing packaging material for dairy products and soft drinks is an interesting issue at the moment. Discussions arise on the costs impacts and environmental impacts of both one way packaging and reusable packaging. The aim of this article is to develop an evaluation tool providing costs and environmental impacts of the PC-bottle and the GT-packs in the dairy sector, considering forward and return flows. The evaluation tool enables the user to analyse the costs and environmental impacts of a supply chain with and without return flows using scenario analyses with respect to the use of various carrier types and the number of return loops. It appears that costs differences between PC-bottles and GT-pack are quite small. The PC bottle has a better environmental profile than the GT-pack. Scenario analysis on the carriers results in the advice to use preferably roll-in-containers with direct delivery, secondly roll-in-containers with delivery via distribution centers, thirdly in case of direct delivery either cartons or crates and cartons in case of delivery via distribution centers.

1 Introduction

This paper focuses on the one way and two way packaging of products in the dairy sector[1]. In 1994 EU regulation on packaging enhanced producers to reduce the amount of packaging waste of various branches of industry (EU 94/62/EC, 1994). Targets of 50-65% of packaging waste stream recovered or recycled should be achieved by the year 2001. Based on this regulation Dutch industry agreed in 1997 (www.minvrom.nl) to target for 65% of packaging material either reused or recycled. New one way packaging material can only be introduced if its environmental impact is less than the impact of comparable reusable packaging material.

Campina Melkunie produces fresh milk both in one way packs and reusable bottles. Given the growing interest in the impact of reusable packaging material on economical and environmental performance, Campina Melkunie wants to gain more insights into the costs and environmental impacts of the supply chain of fresh milk. The problem description is as follows:

[1] This work is part of research carried out recently at the Logistics Center of Expertise of Campina Melkunie B.V.

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

What should be the role of returnable bottles and carriers in Milk Distribution of Campina?

Looking at decision support models available in the literature, we see on the one hand cost models (e.g. Krikke et al., 1999, and Kroon and Vrijens, 1995) and on the other hand environmental (Life Cycle Assessment or analogous) models like (Mekel and Huppes, 1990). The aim of this paper is to describe an evaluation tool that provides the user with cost impacts and environmental impacts of the forward and return flow of a supply chain. The evaluation tool can be used for various scenarios e.g.

- What is the effect of the number of reuse loops on the costs and environmental impacts of reusable packaging material
- What type of carriers is suitable for either one way or reusable packaging material.

Bloemhof et al. (1995) describe a methodology to use environmental information within the decision process of a product mix problem. Using an environmental index it is possible to compare cost-friendly product mixes with environmental friendly mixes with respect to costs and environmental impacts. Bloemhof et al. (1996) attempt to combine life cycle analysis with logistic optimisation while optimising the design of a production network for the pulp and paper industry. Life cycle assessment is used to obtain an environmental performance indicator for each process. These indicators are used in a network flow model to find optimal designs of the pulp and paper network with the lowest environmental impacts. Based on these methodologies the CAMP evaluation tool has been developed. It contains an Activitity Based Costing model combined with a Life Cycle Analysis Tool.

Section 2 describes the company Campina Melkunie. Campina Melkunie produces about 32 brands of milk, cheese, butter, and yoghurts for direct consumer use and industrial products as protein products and lactose products. The sales area contains over 100 countries. Section 3 describes the supply chains for the PC bottle and the GT pack in more detail. In Section 4 we present the evaluation tool CAMP (Choice of Alternative Material Packaging). The CAMP tool is developed to analyse the costs and environmental impacts of the forward and reverse chain of the packaging material of fresh milk. Section 5 deals with sensitivity analysisis and scenario analysis and Section 6 provides our conclusions.

2 Campina Melkunie

Campina Melkunie is an international cooperation aiming at the development, production, sales and distribution of dairy consumer products and ingredients for the pharmaceutical industries. Apart from fresh milk, also cheese and yoghurts are produced with international brand names like Yogho Yogho, Vifit, Yazoo, Joyvalle, Passendale, Milner, Monchou, Tuffi and Landliebe. Industrial products are

sold under the brand names Esprion and Excellion (protein products), Pharmatose (lactose product) and Emser (ingredients).

Campina Melkunie is a cooperation with about 8500 farmers associated. The turnover is 8 billion Euros. The market for Campina contains about 100 countries with a large domestic part in the Netherlands, Germany and Belgium (see Figure 1).

The research focuses on Campina Netherlands, which is a subdivision of Campina Melkunie, mainly producing fresh milk. Production units are in Eindhoven, Hilversum, Maasdam, Rotterdam and Heiloo. These production units also have a distribution center for the delivery of products to buyers in the region. A distribution network between the production units guarantees a full assortment of products in each region. The same networks are used for the collection of reusable packaging material and cargo carriers.

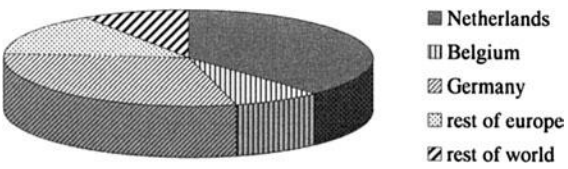


Fig. 1. Spread of Turnover

The mission statement of Campina Melkunie is to add value to milk by (i) being entrepreneurial, (ii) making difference in the chain, (iii) focussing on consumer needs and (iv) caring for people, which results in *"a natural caring for the sustainable values of our nature with an environmental responsibility"*.

3 Fresh Milk Supply Chains

Supply chain management (SCM) or chain integration is an important development in logistics management. SCM is an integrative approach to dealing with the planning and control of the materials flow from suppliers to end users (Ellram, 1991). Companies in a supply chain do not primarily optimise their own activities, but focus on an efficient process management in all parts of the chain. The supply chains of fresh milk do not only contain all processes from suppliers to end users but also the processes from collection to cleaning and refill of the packaging material.

Currently Campina uses both one way and two way packaging for their dairy products. Apart from the traditional package, the Gable Top `GT-pack`, a reuseable plastic bottle, the Polycarbonate `PC bottle` is used. The PC bottle returns af-

ter use whereas the GT pack is disposed of after use. Data considering the costs and the environmental impacts of both the forward chain and the return chain of the bottles can be used to compose a `cheap and green` strategy in the milk distribution.

Besides direct packaging of the milk in bottles Campina uses crates, boxes, crate containers, pallets and roll-in-containers (RICs) for handling and transportation. Except for boxes, all carriers will be returned to Campina for reuse. Campina can choose between different types of carriers, each with accompanying costs and environmental impacts. Next section gives a description of the primary packaging systems whereas Section 3.2 describes the cargo carriers. Section 3.3 focuses on the logistical processes of a PC bottle and Section 3.4 specifically on the return processes in the fresh milk supply chain.

3.1 Bottles

The Gable Top (GT) is a traditional cardboard box used for fresh milk, yogurts, buttermilk and custards. After use, GT ends up in domestic waste. The supply chain of the GT-pack can be described as follows (see Figure 2). Campina buys the packages from suppliers nearby. At the production locations the packs are filled with milk and stapled in crates, boxes or RICs. The crates and boxes are transported on pallets or crate containers to retailers. At the retailer the packs are sold to the consumer and the carriers are returned to Campina. After use the pack ends up as domestic waste. The cardboard box can be recycled or used for energy recovery by incineration.

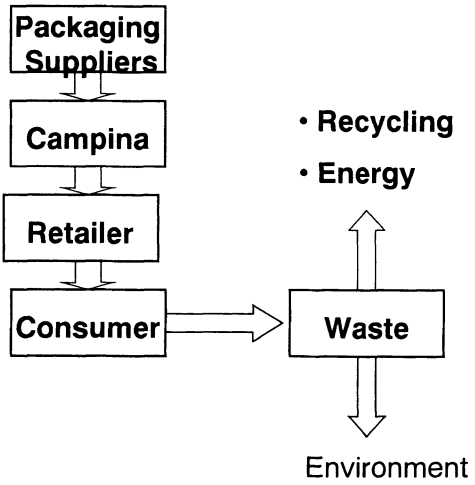


Fig. 2: Supply chain of the GT pack

In 1996, the one-litre Poly Carbonate (PC) bottle was introduced, which is lightweight, recloseable and reusable. At the moment a relatively small amount of

the milk is sold in PC bottles. Campina cleans all returned bottles before refillment and redistribution. A deposit system of one Dutch guilder for a bottle has to prevent bottles ending up in domestic waste. Campina sells refused bottles to the synthetic industry for recycling in dashboards of cars. In practice, a bottle can be used about 27 times before failing the inspection.

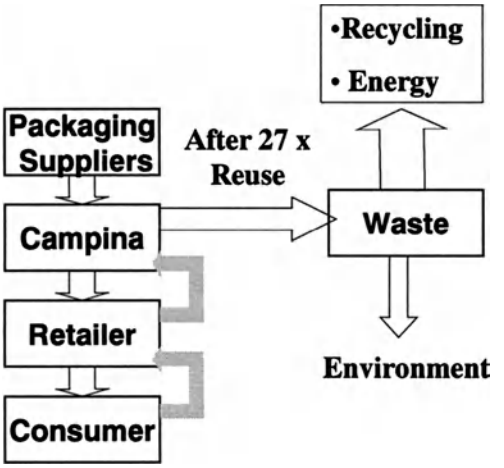


Fig. 3. Supply chain of the PC bottle

Most of the milk is filled up in either a GT pack or a PC bottle. Different packageing forms have also been developed. School milk is packed in small cardboard boxes. Campina collects weekly used packaging at the schools. Collected packaging will be recycled and used for toilet paper and tissues. Fresh milk in a PET bottle is a new product and sold in a 33 cl. format at e.g. fuel stations. The PET bottle is lightweight and reclosable, very suitable for take-away purposes. It is a one-way packaging material that ends up in domestic waste.

3.2 Carriers

Campina Melkunie uses crates, boxes, crate containers, RICs and pallets for handling and transportation. Except for the boxes all cargo carriers must be returned to Campina for reuse.

A *crate* consist of synthetic material. It can hold 20 one-litre GT or PC bottles. Crates can be stacked up on pallets or crate containers. After use the crate will be returned to Campina and reused after testing and cleaning. A drawback of crates is the fact that they use as much space filled on the outward journey as empty on the way back, causing relatively high transportation costs as well as sorting and handling costs of empty crates. Crates have a rather long lifetime and can be recycled afterwards to granulate for new crates.

B*oxes* contain six to twelve one-litre GT packs or PC bottles and are used for some DC-customers. A box can be stacked up on pallets or crate containers. PC

bottles can be stacked to a higher level than GT packs. After receiving and unpacking, the retailer collects the cardboard for recycling purposes.

A *crate container* is a multiple purpose carrier on wheels, used for direct deliveries of crates. Obviously, the crate container cannot be nested as it contains crates.

A *pallet* is mainly used to deliver crates or boxes to distribution centers. Empty pallets can be stapled easily so it requires less space at the return part of the supply chain. In most logistic chains pallets have been standardized to maximise logistical efficiency at both the suppliers and the buyers (the so-called EURO pallets). For fresh milk products Campina uses a return cycle with specific Melkunie pallets.

A Roll in Container is a moveable carrier that is automatically filled with 160 one-litre packs in the production locations of Campina and used as shelf at the retail shops. Using a RIC makes boxes or crates superfluous saving enormous handling costs and time. Its product-homogenity and a high use of shelfspace are drawbacks of a RIC.

3.3 Supply Chain of a PC Bottle

In order to make comparisons of packages and carriers based on costs and environmental impacts a complete description of the fresh milk supply chain is necessary. In Figure 4 we draw a distinction between the forward and the return part of the chain. Figure 4 focuses especially on the logistical process of a PC bottle since the supply chain of the GT bottle has no return part.

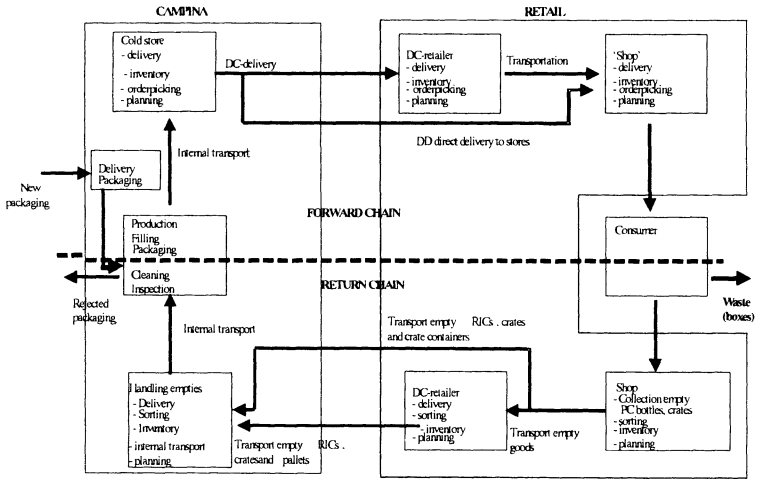


Fig. 4. The logistical process of a PC bottle

Campina deliveres filled PC bottles to distribution centers (DC) and directly to shops (DD). In the return part of the chain the empty bottles are also collected in

crates, using pallets (DC) and crate containers (DD). This reverse process causes a lot of handling and transportion. On average a PC bottle can be reused about 27 times. A *return loop* starts and ends at the process of filling the bottles with milk. After end-of-use the bottle is sold to the synthetic industry for recycling purposes.

In order to satisfy market demand, it is important for each plant location to have enough empty bottles, crates, pallets and crate containers available. It is difficult to forecast the amount of empty packaging and cargo carriers since the empty bottles are often not returned to the original plant. If the inventory of emply bottles is not high enough, new PC bottles have to be bought by Campina. The necessary amount of external supply depends on the inventory of empty bottles, the historical and forecasted supply of filled PC bottles, the average duration and variation of a return loop and the forecasted return rate and failure rate of empty PC bottles. Both new and returned bottles have to be cleaned.

3.4 The Process Flow of a Returned Bottle

The return process of a PC bottle starts when the consumer returns the empty bottle at the retailer. Retailers return crates with empty bottles directly to a Campina production location or via a distribution centre of the retailer following the „full for empty" rule.

At the production location the process flow of the returned bottles is as follows (Figure 5).

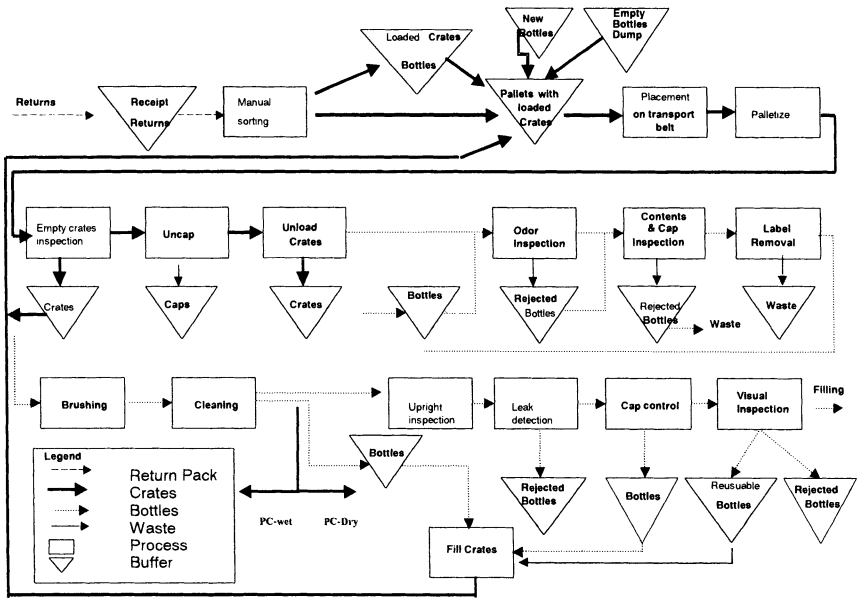


Fig. 5. Process flow of empty PC bottle

After delivery the returned bottles are sorted by hand. Crates full with empty bottles are stapled on pallets and collected at the start of the PC-assembly line. The bottles are uncapped and the caps are collected for external recycling. Then the bottles are put one by one at the assembly line. First the odor check is carried out (is the bottle a milk bottle or not). Then the emptiness control takes place (is the bottle empty) followed by the cap control (is the cap removed succesfully). Next phase is the cleaning phase consisting of removing the labels, brushing the inner side of the bottle and rinse the bottle completely. This process takes about 20 minutes per bottle. After the rinsing process a leackage check is followed by a visual inspection. If the bottle endures the inspection it can be refilled.

4 CAMP

The evaluation tool Choice of Alternative Material Packaging (CAMP) is developed to compare one way and returnable bottles and carriers based on costs and environmental impacts (Kraal, 2000). CAMP is based on the following assumptions:

- All PC bottles contain low fat milk.
- The production and sales quantity of PC bottles and GT packs are the same (in order to compare costs and environmental impacts).
- The PC bottle has on average 27 return loops before end of lifetime.
- Full truck load for delivery of new bottles, caps and labels.
- If a PC bottle is not returned it ends up in domestic waste.
- About 10% of the returned bottles has no cap. These caps are part of domestic waste.
- Incineration of domestic waste takes place in a closed installation.

Table 1 illustrates the various steps in the CAMP tool. Both the cost part and the environmental part consist of three steps. First step is the inventory of the processes and activities within the supply chain. Next step is the determination of the relevant cost drivers and environmental issues. Thirdly, costs and environmental effects are assigned to products. The result is either a cost component or an environmental impact for both the PC bottle and the GT pack.

Table 1. The CAMP evaluation tool

Choice of Alternative Material Packaging	
ABC	LCA
Inventory of activities	Inventory environmental impacts
Determination of cost drivers	Determination of relevant environmental problems
Assignment of cost to products	Assignment weights to problems
COST COMPONENTS	*ENVIRONMENTAL IMPACT*

4.1 Costs of Packaging

The cost part of CAMP is based on Activity Based Costing (Cooper and Kaplan, 1988). The ABC method is used in a dynamic environment with bad predictable demand, short product lifecycles and a broad assortment. The method is based on finding the activities that cause the costs and describe the way they are linked with a product. The ABC method consists of the following steps:

- Inventory of the important activities
- Determine the cost drivers for each activity
- Assign costs of activities to products.

The PC bottle goes through a forward chain and a reverse chain. The GT-pack only has a forward chain but the crates, crate containers and pallets used for the transportation of GT-packs do have a reverse chain.

The total costs can be divided into three cost components:

- Costs of the packaging material itself: These costs including purchasing costs of bottles, labels, caps and glue, transportation costs of the material and a negative cost component of deposit fees;
- Internal costs of the forward chain: These costs include the costs of the filling process, packaging, internal distribution, salaries, energy, distribution from production location to distribution center, distribution from distribution center (DC) to retailer and the activities at the DCs and the retailers.
- Costs of the reverse chain: These costs are both external (activities at the retailer and the DCs, distribution from retailer to DCs and from DCs to Campinas production location and the transportation of waste) and internal (fixed costs for the PC reassembly line, internal distribution, salaries, energy and packaging material).

Table 2 shows the results of the integral cost comparison between PC and GT bottles assuming direct delivery to retail stores with crates on crate containers as carriers.

Table 2: Cost components as percentage of the consumer price (direct delivery)

	PC	GT
Consumer price	1,59	1,25
Packaging	8%	21%
Forward chain	68%	71%
Reverse chain	24%	8%

The difference in costs between one way and returnable packaging systems appears to be limited. The PC bottle has higher costs in the reverse flow but these costs are compensated by low material costs per unit as the bottle can be used about 27 times.

4.2 Environmental Aspects

The environmental impact of the use of one way bottles or reuseable bottles is determined by a Life Cycle Analysis (LCA). According to SETAC (1993) *Life cycle assessment (LCA) aims to evaluate the environmental burdens associated with a product, process or activity by identifying and quantifying energy and material used and wastes released to the environment; to assess the impact of energy and material used and wastes released to the environment and to identify and evaluate opportunities to affect environmental improvements.* LCA can be defined as an input-output analysis of resources or materials and energy requirements in each phase of the life cycle of a product. Usually it is composed of four parts:

- The definition of the scope and the boundaries of the study.
- The inventory quantifying the necessary data in an objective and consistent way using an input-output database.
- The impact assessment classifying the inventory results by environmental indices and their valuation concerning the environmental impact.
- The improvement assessment focussing on the reduction of environmental impacts associated with the system under study.

With the inventory one can identify opportunities for reducing material use, energy requirements or emissions. The impact assessment helps to become aware of the different types of environmental impacts whereas the improvement assessment aims especially in identifying potential reduction strategies. Figure 6 represents the process tree of PC bottles and GT bottles.

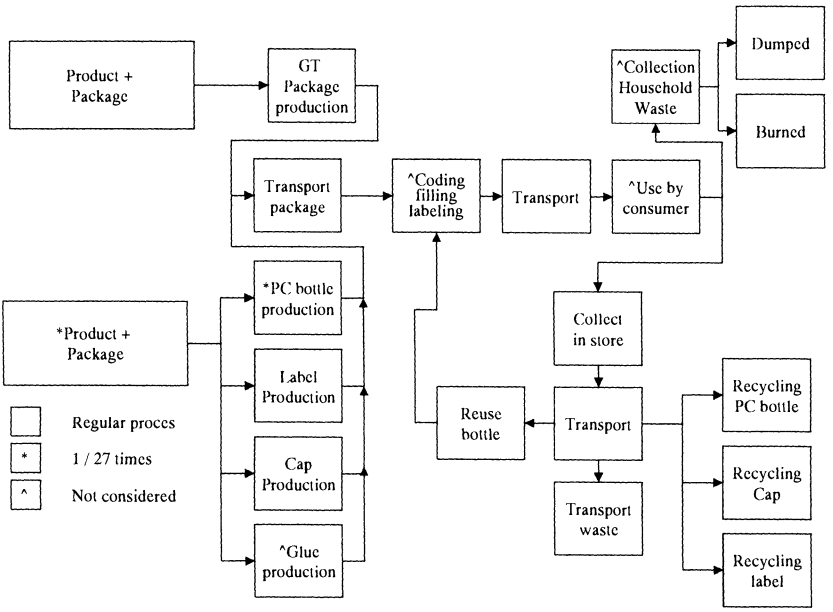


Fig. 6. Process Tree

Assumptions for the environmental part of CAMP are:

- Filling, coding, labelling and using either a PC bottle or a GT bottle makes no difference in environmental impact.
- Domestic waste will be disposed of for 10% and incinerated for 90%.

For each process an ecobalance is made, based on research from Mekel and Huppes (1990). Updating to 1999 has taken place where necessary. The ecobalances are classified based on the contribution to the various environmental problems, resulting in an environmental profile (Table 3).

Table 3. The environmental profile of PC and GT bottles

(in kg/year)	PC bottle	GT bottle
Greenhouse effect	-175	-572
Smog	1.44	2.43
Acidification	7.43	33.5
Nutrification	1.44	2.03
Human Toxicity	11.37	20.85

After normalisation the environmental impact of a PC bottle can be compared to a GT bottle as follows (Table 4)

Table 4. Relative environmental impact of PC and GT bottles

	PC bottle	GT bottle
Greenhouse effect	-1	-3.18
Smog	1	1.69
Acidification	1	4.52
Nutrification	1	1.40
Human Toxicity	1	2.00
Environmental Impact	1	2.82

The Life Cycle Analysis shows that the GT bottle has about three times higher contribution to the environment than the PC bottle.

5 Sensitivity Analysis

The previous sections gave some insight in the costs and environmental aspects of one way and reusable packaging materials in the dairy sector. In order to draw conclusions, it is very important to investigate the sensitivity of the results if some of the input variables change. We give an account of four scenarios.

- Fixed costs at the production location differ with up to 25%. The results of the CAMP model change between 5-7 %, being not decisive.
- Costs of activities in the distribution centre differ with up to 25%. Again results differ with less than 6%.

- Energy use for cleaning PC bottles differs with up to 25%. Environmental impact differs with about 5%.
- Energy use for the production of PC bottles differs with up to 25%. Results of the model change with less than 1%.

Furthermore we carried out sensitivity analysis on *(i)* the number of return loops and *(ii)* the type of carriers.
(i) Changing the assumed number of return loops has a large influence on the purchase costs of PC bottles. If the number of return loops increases, the purchase costs per bottle decrease as well as the costs of buying new bottles. The CAMP model gives the following results: costs of the PC bottle decrease with an increasing number of return loops whereas the environmental impact of the PC bottle slightly increases.
(ii) If the PC bottles and GT packs are stapled in boxes instead of crates, this has the opportunity to get around deposit fee issues. Possible drawbacks are an increased amount of waste and less space in the DC.
Using the RIC instead of crates gives considerable cost savings due to less activities at the retailer and the external distribution. The environmental impacts of RIC and crates in crate containers are about the same. If the PC bottle is transported in RIC, crates are still necessary for returning the PC bottle to Campina.

6 Conclusions

Comparing the costs of the PC bottle and the GT pack in crates gives the following results. Costs for the forward chain are almost the same for PC bottles and GT packs. The return chain for PC bottles is more expensive than for GT packs which is rather obvious. However, the total cost difference based on equal quantities is only limited. The PC bottle has a significantly better impact to the environment than the GT pack.

Given the results of this research the following recommendations hold;

- Increase the sales volume of PC bottles.
- Use a RIC for large volumes of PC bottles and GT packs.
- In case of direct delivery (DD) crates and boxes are equaally attractive.
- In case of delivery via distribution centres (DC) boxes are preferred above crates. Crates however are still necessary for the return chain.

References

Bloemhof-Ruwaard, J.M. / Koudijs, H.G. / Vis, J.C. (1995) Environmental impacts of fat blends: a methodological study combining life cycle analysis, miltiple criteria decision making and linear programming, *Environmental and Resource Economics* 6:371-387.

Bloemhof-Ruwaard, J.M. / Van Wassenhove, L.N. / Gabel, H.L. / Weaver, P.M. (1996), An environmental life cycle optimization model for the european pulp and paper industry, *Omega, International Journal of Management Science* 24:615-629.

Cooper, R. / Kaplan, R.S. (1988): Measure costs right: make the right decisions. *Harvard Business Review* sept/oct: 96-103.

Ellram, L.M. (1991): Supply chain management: the industrial organisation perspective, *International journal of physical distribution and logistics management* 32(1): 13-22.

EU 94/62/EC (1994) European Parliament and Council Directive of 20 December 1994 on packaging and packaging waste, *Official Journal L365*, 31-12-1994:0010-0023.

Horngren, C.T. / Foster, G. / Datar, S.M. (1996) *Cost Accounting*, London: Prentice-Hall International.

Kraal, A. (2000) ***PAK (e)en FLES,*** master thesis Faculty of Business Administration, Erasmus University Rotterdam.

Krikke, H.R. / Van Harten, A. / Schuur, P.C. (1999), Business case Oce: reverse logistic network re-design for copiers, *OR Spektrum* 21-3:381-409.

Kroon, L. / Vrijens, G. (1995) Returnable containers: an example of reverse logistics. *International journal of physical distribution and logistics management* 25(2): 56-68.

Mekel, O.C.L. / Huppes, G. (1990) Environmental effects of different package systems for fresh milk, Leiden, Center of Environmental Studies.

SETAC (1993) Guidelines for life-cycle assessment – a code of practice, Brussels.

www.minvrom.nl

Chapter 3

Distribution Logistics and E-Commerce

The Logistics Behind the Enter Click

René B.M. de Koster

Rotterdam School of Management, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, Netherlands

Abstract. Internet gives powerful opportunities to retailers to boost sales, increase market share and generate new business through new services. One of the challenging questions that retailers are faced with is how to organise the logistic fulfilment processes during and after the transaction has taken place. This article gives an overview of the different decisions that have to be taken with respect to the distribution of products to consumers. The different alternatives for the distribution are discussed and a model is presented with relations between the company's objectives and characteristics and choices in the distribution. Some preliminary research into such relations is discussed. It may be concluded that there are ample opportunities for further research.

1 Introduction

Internet is still becoming increasingly important as a new sales channel. Although the total value of Business-to-Business (BtB) e-commerce is much larger than the total value of Business-to-Consumer (BtC) e-commerce transactions (see f.e. several reports of Forrester Research), the number of BtC e-commerce transactions increases fast. Also, more and more companies start setting up web pages aimed at reaching consumers and more and more consumers have experience in buying products via the internet.

Currently, the internet is only responsible for a minor segment of the $2.2 trillion retail market. This segment is however, projected to grow fifty times faster than in-store shopping (Palmer, 2000). In May 2001, The Gartner Group stated that consumer purchases via the internet were worth $20 billion in 1999 and predicted a rise to $147 billion retail in 2003 (Ferrara et al., 2000). The predictions of the different marketing research organisations are, however, far apart. One year earlier, Forrester Research, for example, estimated internet sales much lower and predicted a world-wide turnover of $3.2 billion via the internet in 2003 (McCullough, 1999). What they have in common is the prediction of the fast growth of BtC e-commerce transactions. According to a Taylor Nelson Sofres study of 32,000 people in 27 industrialised countries, about 27% of the people are online. 10% of the net surfers shop online in a month (The Industry Standard, 2000).

The number of companies with a web page has increased rapidly as well. Already in 1997 all, but nine, top 100 retail companies in the USA had their own homepage (Morganosky, 1997). These web sites were mainly used for providing information on the company. Orders could be placed at only one-sixth of these

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

web sites. This percentage has increased considerably. Four types of companies that sell online to consumers can be distinguished.

- Product manufacturers, such as Dell (computers), Unilever (cosmetics, products with high added value), Numico (food additives), BOL (books, media). Nevertheless, direct sales to customers are still not very common for manufacturers.
- Traditional retailers and wholesalers, such as Barnes & Noble, Albert Heijn, Tesco, Makro, Karstadt, Quelle.
- New internet companies without physical assets (intermediaries), such as Let's-buyit, Boxatwork, E-bay. Until 1999 the number of companies of this type has been booming. In 2000 there has been a rapid decline (Boo, Etoys).
- New internet companies, with physical assets (for example stock, warehouses, or trucks), such as Amazon, Peapod, Hotorange, Maxfoodmarkets.

Buying products via the internet may bring considerable advantages for customers. These advantages may hold for digital as well as for non-digital products. Advantages for buyers and or sellers are given in Table 1 (see also Kambil et al., 1999 and Kern et al., 2000).

Table 1. Advantages to buyers and sellers of online shopping.

Advantages for sellers	Advantages for buyers
• Accessibility. Especially in western countries internet has become available in many households. This potentially can increase the market for sellers. • Information can be given to customers without delay. Examples are frequently asked questions, price changes, special offers. • Novelties can immediately go to the customers, such as new or test products. • Possibility for customisation of service and information. By data mining historic purchase behaviour of customers, sellers can use this to give customers alternatives, special discounts or use standard customer-specific shopping lists that are easily modifiable. • Potential for new business models. Think of electronic markets (E-bay, Let'sbuyit).	• Ease. Internet orders are usually brought home. • Time saving in searching, and actual purchasing (Doherty et al., 1999, Morganosky and Cude, 2000). Customers can save time compared to in-store shopping. • Improved transparency. Buyers can compare prices and services of different sellers. • Improved pricing, due to increased market competition and economies of scale available to sellers from aggregating demand (Doherty et al., 1999, Morganosky and Cude, 2000). Lower prices may also be due to skipping services of intermediaries (Benjamin and Wigand, 1995, Grover and Segars, 1999, Sakar et al., 1995) or, in case of international shipments, avoidance of import duties and value added taxes.
• Availability. The internet is available around the clock, at times convenient to the customer.	

In spite of the above hypotheses and findings (partly based on field research (Doherty et al., 1999, Morganosky and Cude, 2000), recent research has demonstrated that these potential advantages for internet customers do not necessarily all come true. In longitudinal research on purchasing prices and purchasing lead time for 112 durable consumer products via the internet, it was found that product prices and purchase time do not significantly differ from in-store prices and buying time (Palmer, 2000). Other researchers suggest that although certain supply

chain intermediaries may be eliminated, there still is a need for certain intermediary functions and parties ('cybermediaries', Sakar et al., 1995, Jin and Robey, 1999). Therefore, the realisation of lower prices need not necessarily come true.

One of the major problems for companies selling non-digital products via the internet to customers is to deliver the goods and thereby meeting the customers' expectations. In research carried out by European consumers' organisations it was found that logistics aspects, such as delivery lead times (or delivery at all), or simply meeting promises were not met by a substantial part of the investigated internet companies (Consumentengids, 1999).

The problem areas for e-tailers may be summarised as follows:

- Channel conflicts. Several manufacturers and wholesalers have made attempts to sell products to customers, thereby bypassing intermediaries (Unilever: supermarket chains, Levi's: dealers, Compaq: dealers). Most of them have given up quite rapidly for several reasons. It is not easy to bypass existing distribution channels, as the intermediary companies involved may refuse to support the product any longer.
- Difficulty of keeping customers. It may be difficult to attract customers in the first place, but it is probably even more difficult to keep them. When the price/quality offer of both the product and the service rendered are not right, customers will not return.
- Difficulty of handling returns. In several EU countries, mail order companies are forced, by law, to accept returns at no costs for the customer, within a certain period after the buy. This also holds for online purchases. In mail order business returns may amount 30% of the total sales, De Koster, et.al. (2001) depending on the product type (fashion is the worst). In handling the returns, problem areas are: customer service options, speed of credit, willingness to compensate return shipment costs.
- Untimely deliveries. E-tailers often fail to deliver in time. This is especially true for e-tailers without physical assets. Since they do not own the stock, they often have insufficient insight in the saleable stock level, which leads to unkeepable promises and misexpectations, with respect to delivery times.
- International shipping and home delivery is either expensive (express carriers) or relatively slow with unwarranted service (mail).
- Handling large volumes with a virtual organisation. When volumes become large, it becomes difficult to keep outsourcing the physical process, since the company may loose control.

As customer wishes may be beyond what is possible or desirable for the supplying company, the need for a good business model becomes evident.
Companies with unique business models that are specific to the Web, like Priceline and Ebay, are considered possible winners, while those with great customer service and new revenue streams like Amazon are also well-positioned. But across the board, companies will have to show a clear path to profitability and the ability to manage costs. In another Forrester report (see Jedd, 2000) the world wide number of parcels shipped daily by e-tailers was estimated at 650,000. By 2003 this

number could grow to 4.2 million daily shipments. The fulfilment problems are therefore, not expected to lessen.

In this article, the different logistic business models possible will be assessed. First an overview will be given of the different decisions with respect to distribution that have to be taken. We focus on decisions such as the choice of distribution channel, use of warehouses, delivery area selection, and degree of outsourcing. Then a number of distribution models will be compared in a qualitative manner. The independent variables that influence the ultimate choice for a particular model will be discussed. Also, the impact of the choices made on logistic costs is treated. Previous research carried out in this area, both empirical and model-based is discussed in section 5. Finally, some conclusions are drawn.

2 Decisions for BtC E-Commerce Distribution

The fulfilment model determines the way the orders are fulfilled, the number and type of facilities used in this fulfilment process, the area where customers will be delivered and which processes will be kept in-house and which will be outsourced. The main decisions that have to be taken are summarised in Table 2.

The decisions in Table 2 are somewhat interrelated. For example, in case delivery is made world-wide, it is very unlikely that the internet company insources all the transportation. Also, if the delivery area is large, transport will usually not be organised from a store. There is also a natural dependency between the strategic and the tactical decisions. If products are not owned, they can not be supplied from own stock.

The decisions are for a large part identical to the sort of decisions that are taken by old-economy retailers. There are, however, some differences. For example at strategic level, a retailer with existing suppliers, logistic service providers, stores and warehouses may or may not choose to use these existing services and infrastructure. If he does so, the limitations are then in the delivery area, the assortment, packing of products and the home delivery aspects. The question is, whether these existing service providers and facilities are able to carry out the new activities, which potentially may grow very rapidly. Other differences can be found at the tactical level in the combined setting of delivery times, minimum order quantities, delivery fees and delivery time windows for home delivery. Also, the delivery area may deliberately be restricted to keep costs under control. There are not many old-economy companies that have faced the impact of customer oriented delivery models that some e-tailers have (or had) come up with. For example, Webvan before it went bankrupt in 2001, had delivery time windows for fresh food of 30 min in urban areas around large cities in the USA. Streamline (recently taken over in part by Peapod), even installed a secure refrigerator in the customer's garage, to be able to deliver products, without the customer's presence. EGuo.com guarantees delivery lead times in Beijing for a fairly large assortment of 1 hour or less (Xie and Wang, 2001). Such ultimate customer orientation is unprecedented for traditional retailers. See also Småros et al. (2001) and Punakivi and Holmström

Table 2. Strategic and tactical decisions that have to be taken about the fulfilment structure

Strategic decision level
• Should stock keeping intermediaries be used in the distribution? For manufacturers, this may be a logistic service provider, for retail chains this may be stores. If so, which intermediaries: how many stores / warehouses and where? • Which market segment and which assortment should be aimed at? • Where and how should this assortment be bought (quantum discounts)? • Which products should be kept on stock (at own account), which should be bought on order? • In which area should be delivered? • What should the distribution structure be: direct from the stock-keeping warehouse or via hub-and-spoke? • What should the return strategy and related service proposal be? • Should the warehousing operation be outsourced? • Should transport and delivery be outsourced? To what extent? • What delivery model should be chosen: home delivery, fixed drop-off point, or customer specific drop-off point?
Tactical decision level
• To which companies should the fulfilment be outsourced? • Which delivery times should be offered? • Which products / orders should be supplied from stock, which delivered directly by suppliers and which should be cross-docked? • Which delivery time windows should be offered? Can the customer choose a window? • Must the customer be present at delivery? • Which part of the fulfilment costs can (or should be) charged to which customer? • Minimum order sizes?

(2001), who give some additional examples how new services can be implemented in combination with home delivery.

3 Fulfilment Strategies

One can observe different choices for fulfilment structures, delivery regions, outsourcing strategies. The fulfilment structure can take the following forms (see also De Koster, 2001a and Kämäräinen et al., 2001).

- Distribution from existing stores
- Distribution from existing store distribution centres (DC), i.e. distribution centres that supply conventional stores
- Distribution from special DCs for internet only customers
- Hybrid structures, using the different facility types mentioned above.

All distribution structures are feasible for both old-economy and new-economy retailers. Peapod in the USA and Boxatwork in The Netherlands are examples of new-economy food retailers that operate as an intermediary and supply customers

via existing stores and/or DCs (although Peapod is gradually moving towards an asset-based business model). The option to supply from an existing store DC (for retail chains) is only feasible in case of a limited number of internet orders. Bruna (books, media), Freerecordshop (media) are some examples in The Netherlands. In such a store DC, products are usually stored in pallet racks, with long travel distances per order. These long travel distances are justified since order sizes are large (for example, three or four roll cages or pallets in one round trip, where each roll cage represents a delivery to one store). However, the orders of internet customers are usually small. This means that nearly the same distance (time) has to be travelled for a fraction of the volume of a store order. Or, if multiple customer orders are picked in one route, a complex additional sorting process is necessary. In both cases the customer orders have to be carefully packed after the order picking, a process new in the store DC. Also information systems, inventory management and the transport system from the store DC, have not been designed for these many, small customer orders. In conclusion, we can say that order picking in the store DC for internet customer orders usually is not a good solution for the fulfilment problem.

The two remaining main options, fulfilment from existing stores or from a special internet warehouse both have their pros and cons. Order picking from a store for internet customers, packing the orders and home deliver them, is not a process for which a store has been designed. In general, the layout and product-to-location assignment in a store is such that relay times of customers are maximised (think of a supermarket or department store). This can be achieved, for example, by storing fast moving products relatively far away from each other and having aisles with a minimum of possibility to make short cuts, to maximise the exposure of products, leading to impulse buys. Also, products are not stored in the shelves sorted on unit turnover within a product family, but often on product margin. Products stored at eye-level are the products with high margins. Furthermore, order pickers working on internet customer orders in a store may disturb ordinary customers (who takes the last remaining product?). This situation is slightly different for stores with an adjacent warehouse. In such situations, the order picking process can take place in this warehouse, thereby avoiding the previously mentioned problems. However, especially in densely populated areas, one can see that storage space is increasingly sacrificed for sales area. In conclusion, the efficiency of order picking for internet customers in stores is low, leading to relatively high labour costs per order.

The best (i.e. most efficient) solution is probably to have a specially designed warehouse, designed for internet customers only. This facility can be designed for picking many small customer orders by order, directly into the different packaging types (boxes). Internal travel times can be minimised by using appropriate systems, such as carousels (Webvan), case-flow racks (Albert Heijn: ah.nl), pick-to-light systems (Centraal Boekhuis, fulfilling orders of BOL), sorters for sortation on customer order (Wehkamp) and appropriate information systems. De Koster (2001b) gives an overview of storage and handling systems that can be used in such internet warehouses. Kämäräinen et al. (2001) give some insight in the different order picking systems that can be used. The transport to the customers can

be organised from the centre. However, setting up an internet DC with such internal systems is expensive and only justified in case of sufficient scale. Another disadvantage for companies with stores may lie in a high density of this store network. In Europe, some retail chains have such a dense store network. Direct delivery from an internet DC to customers living in the service area of an existing store can cannibalise the local sales, thereby affecting turnover and profits. Especially franchisees may oppose to such strategies. Possibilities to cope with such unwanted effects are to supply internet customers only in regions with low store density, or to focus on different market segments. Ahold is an example of a company focussing on different target groups (Ahold, annual report 1999). It has planned to roll out national and even international internet supermarket concepts for all its food chains (over 30), but current internet product prices are substantially higher than in the store.

Table 3. Advantages and disadvantages of different fulfilment strategies.

	Strenghts	**Weaknesses**
Order picking and delivery from local stores	- low investments, easy to set-up for companies with stores - fast response times - franchisers can be involved - knowledge of customers and market	- limited (unequal) assortment - stores not designed for efficient order picking - additional processes necessary - interference with existing customers - capacity limitations - inefficient transport
	Opportunities	*Threats*
	- rapid expansion possible - new customers can be acquired	- limited extension of assortment - limited growth - cannibalisation of own market
Order picking from internet DC	*Strenghts*	*Weaknesses*
	- assured quality - large capacity possible - own assortment - layout, design fit for small-orders picking - economies of scale obtainable - efficient transport - larger area can be served than from stores	- high investments (depending on degree of mechanisation) - volume is needed - little knowledge customers and market - assortment has to be built up - long transport distances; time windows have to be met
	Opportunities	*Threats*
	- economies of scale	- inexpensive labour is scarce - franchisers opposition - cannibalisation own market

The advantages and disadvantages of the different fulfilment strategies are summarised in Table 3.

4 Variables Determining Distribution Strategy

The choices with respect to the distribution strategy that have been introduced in section 2, depend on internal and external objectives of the company (Van Goor et al, 1996). In this section these choices (mostly the strategic ones) are translated in a number of dependent variables. The e-commerce distribution strategy of a company can be characterised by:

- fulfilment facility type (via internet warehouses, stores, hybrid or store warehouses),
- the position of the main distribution customer order decoupling point (CODP); this is the stock point in the distribution chain where the items are stored. It can be at retailer, wholesaler or manufacturer level (Van Goor et al., 1996),
- the fulfilment centre density. This is the number of stores or warehouses per square kilometre, which varies per region,
- the delivery area size. Companies may fulfil regional orders only, or they may deliver in a whole country, or even world-wide,
- degree of operational flexibility. This is the ability to adapt to late customer orders, or to late changes in orders,
- used transportation mode. Transportation modes can be air, train, ship, truck or van, bike,
- number of transport vehicles used in delivery,
- product return rate,
- number of delivery points. Delivery can be at the customer's home address, customer's employer, or a pick-up point. The number of points will vary per option,
- type of delivery: attended or not attended
- degree of operational outsourcing,
- fulfilment costs per order.

The choice for these variables as being dependent, is to some degree arbitrary. For some of the variables (such as costs, or product return rates) companies obviously have objectives. We see them as dependent in the sense that they can be established objectively and depend on the company's policy and external objectives. The independent variables are either market determined (order volume) or determined by the company's objectives.

The dependent variables are not fully independent of each other. For example, the number of vehicles used will depend on the delivery area size. The choice for the fulfilment centre type will depend on where the stock is stored in the distribution chain. It is expected that the order costs will depend on nearly all other dependent variables. The a priori expected relations between these variables have been indicated in Table 4. In this table, for numeric values, a "+" or "-" stand for a

positive, or negative, correlation respectively, between the independent and the dependent variable. In the case of non-numeric variables, a "+" stands for an expected significant association.

Table 4. Expected relations between independent variables and the consequences for distribution decisions.

Expected Influence of / *on...*	Order fulfilment costs (+ is higher)										
	+	+	+	+	+	+	+	+		+	+
	Fulfilment facility type	Position of CODP in chain	Fulfilment center density (+ is higher)	Delivery area size (+ is larger)	Degree of operational flexibility (+ is higher)	Used transportation mode	Number of transport vehicles (+ is larger)	Product return rate (+ is higher)	Type of delivery	Number of delivery points	Degree of operational outsourcing
Organisational characteristics											
Type of organisation (internet based or traditional)	+			+							+
Product characteristics											
Value density		+	+								
Length of product life cycle			+					+			
Assortment type	+	+	+	-		+		+	+		-
Assortment width	+	+									-
Environmental characteristics											
Availability of LSPs	+		+	+						+	+
Market characteristics											
Order volume	+	+	+		-						-
Market density	+		+							+	-
External objectives											
Order fill rate (required)			+		+						
Delivery time (promised)				+			-	+			
Delivery time window size (promised)			-	+			-				+

The relations assumed in Table 4 have not thoroughly been investigated by empirical research. Research in this area is, in general, difficult since the business is still very immature. This leads to choices for adaptation of certain business models that later are abandoned easily. Also, companies rise and fall very fast and the data of the companies also changes rapidly. In fact, it is hard to find processes that are more or less stable over a few months in BtC e-commerce companies. A research overview is given in section 5.

The costs per order play a significant role in the choices made and will depend strongly on the choices made in distribution, as illustrated in Table 5 and Table 6. The three companies in Table 5 have (or rather "had", since Streamline no longer exists, Webvan is on the edge of bankcruptcy and Peapod changed its structure since) different distribution structures. Webvan has a hub-and-spoke model and distributes from a central warehouse in major agglomerations. Streamline delivers from a central warehouse and Peapod uses associated stores. These structures bring different costs and resulting gross profits per order. Peapod apparently looses money on every order, but also Webvan and Streamline loose money as the gross profits have to cover the substantial fixed terminal costs (a warehouse of Webvan costs $35 million, including the transhipment centres).

Table 5. Costs for three different e-tailers, based on, respectively confidential McKinsey report (1999), Macht (1996), www.peapod.com (mid 1999).

	Webvan	**Streamline**	**Peapod**
Sales + additional income	100.0%	100.0%	100.0%
Purchase price	73.4%	72.0%	75.0%
Logistic costs	13.5%	19.0%	24.0%
Overhead	2.5%	3.05	15.7%
Marketing	2.2%	?	10.5%
Other costs	1.5%	-	-
Gross profit/order	6.9%	6.0%	-25.2%

Table 6. Cost comparison (in NLG) of four different fulfilment options in food distribution. Source: Van der Laan (1999).

Activity	**In-store shopping**	**Online from store**	**Online direct from warehouse**	**Online via satellite station**
Warehouse costs	3.90	3.90	12.50	14.00
Transport warehouse to store	0.90	0.90	-	0.80
Customer order costs	-	3.20	3.00	3.40
Store + payment costs	8.10	13.10	1.20	1.50
Transport to home	-	5.00	8.50	5.50
Return handling	1.20	2.00	1.80	2.30
Customer service	0.90	1.90	2.50	3.00
Total	15.00	30.00	29.50	30.50

Table 6 compares the delivery costs for three different online distribution options (online from the store, online direct from warehouse and online via satellite station) with the base case where the customer shops in the store. It appears that the majority of the costs is in handling (in the store or in the warehouse) and in transportation. Especially the home delivery part of the transport is expensive. Next to the high costs of home delivery as compared to the traditional fulfilment model, where the customer picks up his/her products, a second problem is that for new economy companies the gross margins are often very small, due to lack of buying power. If, for an average order of NLG 100, the gross margin is less than 30%, all online options in Table 6 will lead to losses. Kämäräinen et al. (2001) give some insight how costs can be cut in warehouses by applying different order picking strategies. Punakivi and Saranen (2001) give insight in costs per stop and per route for home delivery of food based on data of 1639 orders of 1450 households in the Helsinki area.

Table 6 assumes that for every option the delivery area is a given. In general, a warehouse can serve a larger area than an online store. The costs per order will also depend on the delivery area size. For both distribution from a store or a warehouse, there is an optimum size of the delivery area that minimises costs. See section 5 for a further discussion.

We will now discuss some of the independent variables and their expected impact on the dependent variables.

Type of organisation

Old-economy companies often already have an existing distribution channel (stores, retailers, dealers) that distribute products for them. If such companies start using internet as a new sales channel, they can, in many cases, use the existing distribution channel. That is, provided the delivery area and the assortment do not change. Advantages of using existing stores can be found in Table 3. Tesco in the UK uses its stores also for internet order fulfilment. Albert Heijn in The Netherlands carefully selects stores in certain areas that deliver internet orders; in some areas with a high customer density, fulfilment is done from a warehouse (De Koster and Neuteboom, 2001).

Value density and product life cycle

According to Van Goor et al. (1996) value density has an impact on where the CODP in the chain will be positioned. It is preferred to store stock of highly valued products upstream in the chain to reduce inventory carrying costs and risks involved in stock keeping. A similar statement holds for products with short life cycles.

Assortment type

Assortments offered to the consumer can be simple or complex. "Complex" products are products that are complex with respect to storage, handling, transportation or delivery. When the assortment consists for a significant part of such complex products, it is considered to be complex. Food (fresh and frozen) and

washing machines (heavy, large and needing installation at delivery) are such complex products. Complex assortments immediately impact the distribution structure (De Koster, 2001a). Simple assortments can be delivered without any difficulties to any part of the world, at least in principle. Complex products can not so easily be handled and shipped over long distances, certainly not at low rates. Fresh products have to be chilled in transport and storage which also limits transport distances. Outsourcing the storage and delivery of such complex products is also not easy, since special equipment (conditioned trucks and warehouses) and special skills and certificates (like HACCP) are required.

Assortment width

The wider the product assortment, the more difficult the management process becomes, thereby impacting for example the outsourcing strategy. All products have to be maintained in the warehouse information system, they have to be visible on the web site and on-line availability may have to be displayed. If multiple facilities deliver the products, then one must be sure that sufficient stock of every product is available in any of these facilities. It may even be necessary to show an internet customer the actual stock of the product in the facility from which (s)he is delivered, in order to prevent out-of-stock situations. This means that inventories of wide assortments will be kept upstream rather than in every store.

Availability of logistic service providers

Transportation companies with a dense world-wide delivery network, with knowledge of international import duty and tax rates that can deliver fast, reliably and at a low price, are non-existent. The result is, that the companies that do ship globally to internet customers, usually only ship to a few countries in Europe, Asia and America where they can fill orders from local warehouses (McCullough, 1999). The availability of good logistic service providers that have knowledge of handling and storing the assortment and that can meet the other service objectives of the company will also impact the opportunities for outsourcing.

Order volume

The number of weekly orders can be small, medium or large. McCullough (1999) makes a split of up to about 100 orders a week, up to about 10,000 orders per week and larger than 10,000 orders per week. According to this report, this number of orders determines whether the company should outsource the warehouse operations or not. The groups with the smallest and largest number of orders should keep (or take) operations in house, the medium group should outsource the operations.

The number of weekly orders will also impact the fulfilment facility type and number, since a large number of small internet orders can not easily be handled by existing facilities that are designed for different purposes. Large volumes in general lower the degree of operational flexibility (see, f.e. Bolwijn and Kumpe, 1998).

Market density

The market density is the distribution of customers over the delivery area. The distribution of customers will have impact on the choice for a distribution centre or fulfilment from a store and also on the fulfilment centre density.

Order fill rate

If order fill rates must be high, it is necessary to either have sufficient stock in all fulfilment centres, or to frequently replenish them. It is difficult to keep sufficient stock on many locations, therefore order fill rate will have some impact on the fulfilment centre density.

Delivery lead-time

The shorter the lead-time offered, the more service is offered to customers. Often, the customer can select delivery lead times himself, where shorter delivery lead times are more expensive. The delivery lead-time has impact on the delivery area. The longer the lead-time, the larger the area that can be served at low costs. Short delivery lead times can only be realised in case either the delivery area is small, or if the additional costs of rapid shipments can be charged to the customer. This can only be done in case of special, rather expensive products. Returns will also be impacted by delivery lead-time. This is particularly true for gift items and impulse buys. If a customer buys a product as a gift, (s)he expects it to arrive in time for the occasion. If not, the product is simply returned.

Delivery time-window size

The wider the time-window, the more easy delivery becomes. If time-windows are very narrow, then more delivery vehicles are needed and, in order to make sure that the delivery is in time, the delivery area size will become smaller.

5 Research on Distribution Strategies

In this section, an overview is given of some research that has been carried out with respect to the research model sketched in the previous section. Research is mostly of an exploratory nature (see, f.e., Tanskanen, 2001, Småros et al., 2001) or it gives an overview of different solutions for fulfilment (De Koster and Neuteboom, 2001, Kaipia, 2000, Pflaum et al., 2000, Pyke et al, 2001). De Koster (2001b) explicitly treats the storage and material handling systems appropriate for e-commerce fulfillment centres. Quantitative empirical or modeling research is still scarce. In De Koster (2001a) some of the relations as sketched in Table 4 have been investigated for 36 companies that were active in the BtC e-commerce sector, during at least six consecutive months. The companies stem from seven different countries.

The results are that there are some positive associations, which are summarised in Table 7. From this table, we can conclude that complex-product operations are usually not outsourced. Organisations that traditionally work with stores use these to a large extent in e-commerce fulfilment as well. Short delivery lead times lead to relatively small delivery areas. Assortment width appears to be not significant for any of the tested dependent variables.

Table 7. Chi-square tests for association, with (exact) significance levels (2-sided test) between parentheses and number of observations between brackets. N.S.: not significant.

	Fulfilment facility type	Operational outsourcing	Delivery area
Product complexity	N.S. [36]	14.5 (0.000) [24]	10.4 (0.004) [26]
Assortment width	N.S. [36]	N.S. [24]	N.S. [26]
Order volume	3.13 (0.12) [29]	N.S. [22]	4.89 (0.040) [25]
Delivery time	N.S. [26]	3.96 (0.074) [22]	13.54 (0.001) [25]
Type of organisation	11.06 (0.001) [36]	8.87 (0.005) [24]	3.71 (0.105) [26]

Ranchhod and Gurãu (1999) have carried out research that is somewhat related. They focus on direct (online) and indirect marketing channels as dependent on characteristics of the company (size), the clients (business or consumers) and traded products (digital or non-digital, unit value and unit volume). They show that the implementation of a particular internet-enabled distribution strategy (online or via intermediaries) depends on these characteristics.

Kämäräinen et al. (2000) and Punakivi and Saranen (2001) investigate the impact of variables similar to the dependent and independent variables in Table 4 on the delivery costs per order and the total mileage needed to deliver a given set of orders in a given area. The results are obtained via simulation, using routing software from CAPS Logistics. They investigate scenarios with different delivery times and delivery time window sizes, as well as attended and unattended goods reception.

As yet there is not much research in quantitative modelling. An interesting subject might be to model the best delivery area size. Although it may seem attractive to draw customers from a wide area, delivering in this area may be rather complicated. Old-economy companies may want to use their existing facilities (stores, warehouses) and forwarding companies. These facilities and companies can not always easily switch to new delivery regions. New regions may require new delivery codes in the information system, new sorting lanes on the sorter machine, new shipping label types, documentation in other languages etc. According to the research of Forrester (McCullough, 1999), three-quarters of the interviewed US retail firms are unable to register international addresses accurately or price total delivery costs. Also, the forwarding company may not deliver in the particular region. The transaction costs related to the switchover to the new regions may be greater as the total new region becomes larger.

In Fig. 1 the relation is sketched between the size of the delivery area and the distribution cost per order, for distribution from a (new) warehouse as compared to distribution from an existing store, for two different demand densities (high and

low). The graph is based on work of Daganzo (1991) and Daganzo and Errera (1999). The distribution costs consist of facility costs, which have been prorated per delivery cycle and handling and transportation costs, which have been drawn from Table 6. This table shows the trade-off between transportation and handling costs for the two types of fulfilment options. For small delivery areas, the order distribution costs are proportional to the reciproke of the delivery area size. For large delivery areas, these costs are proportional to the square root of the delivery area size. From the graphs, one can see that the optimum delivery area size of a store is smaller than that of a warehouse, the optimum delivery area size decreases as the demand density increases and for large delivery areas, using a warehouse leads to lower costs than using a store. In these graphs, it is assumed that vehicles have a fixed capacity (larger for warehouses than for stores), but the facility handling capacity (which is much smaller for stores than for warehouses) has not been included.

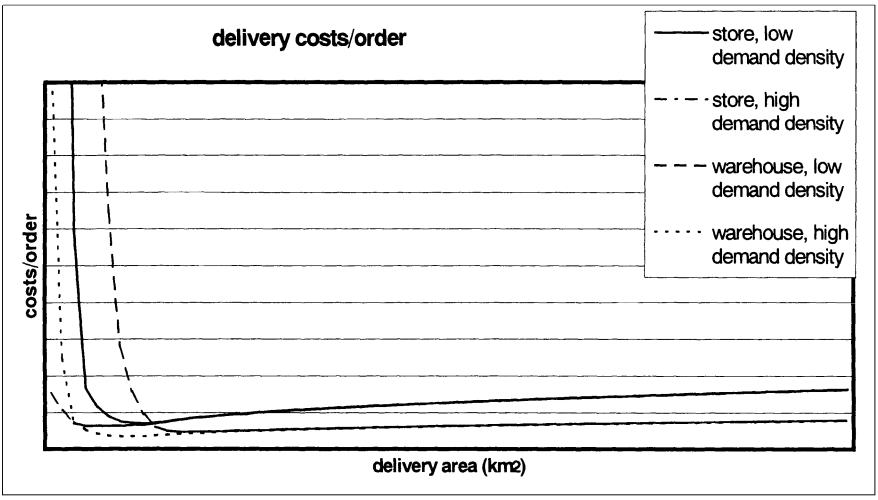


Fig. 1. Total delivery costs as a function of the delivery area size

In view of the high fulfilment costs, there certainly is a need for models that help in deciding the best delivery lead times, delivery time windows for certain product-market combinations, or the best delivery point option (at home or at pick-up points, either or not attended). Daduna (2001) investigated the impact on vehicle routing when the time windows are not tight, but more flexible. Cattani and Souza (2001) optimise the profit, under different stock operating policies, when a certain part of the stock can be reserved for customers that are willing to pay a higher fee for rapid delivery.

Another interesting area is the determination of prices that can be charged for certain services. However in this area, still much research has to be done.

6 Conclusions

In this paper we have given an overview of important decisions with regard to the distribution structure of BtC e-commerce companies. Basically, distribution can be carried out from a special internet warehouse, from stores or through some hybrid structure. All solutions have advantages and disadvantages and have also cost implications. Many companies have to make decisions with respect to the choice of service level that has to be offered to customers (like choice between delivery option, delivery speed and delivery time accuracy), the delivery area that will be served, the assortment that will be offered and the costs which can be charged to such services.

Although we have shown some preliminary results, it is clear that both quantitative empirical and modelling research are needed to bring insight into these problems.

References

Ahold, (1999) Annual Report

Benjamin, R., R. Wigand (1995), Electronic markets and virtual value chains on the information superhighway, *Sloan Management Review*, Winter, p.62-72.

Bolwijn, P.T., T. Kumpe (1998), *Marktgericht ondernemen*, Van Gorcum, Assen.

Cattani, K, G. Souza (2001), Inventory nuances for e-commerce companies, Proceedings of the 12th annual conference of the Production and Operations Management Society.

Consumentengids (1999), Purchasing on the internet (1) (in Dutch), Vol.47, Nr.8, 1999, p.56-58.

Daduna, J.R. (2001), Tourenplanung mit unscharf definierten Lieferzeitpunkten für die Abwicklung von Kundenaufträgen beim Electronic Shopping im Lebensmitteleinzelhandel, working paper, Fachhochschule für Wirtschaft, Berlin.

Daganzo, C.F., A.L. Erera (1999), On planning and design of logistics systems for uncertain environments, in: M. Grazia Speranza and P. Stähly (eds.) *New Trends in Distribution Logistics*, 3-21.

Daganzo, C.F. (1991), *Logistics systems analysis*, Springer Verlag, Berlin.

De Koster, M.B.M. (2001a), Distribution strategies for e-tailers, working paper, Rotterdam School of Management, Erasmus University Rotterdam.

De Koster, M.B.M. (2001b), De logistiek achter de "Enter"-toets, in: J.P. Duijker et al. (eds.), *Praktijkboek Magazijenen en Distributiecentra*, Kluwer, Deventer, 1.4A.01-1.4A.12.

De Koster, M.B.M., J. Neuteboom (2001), *The logistics of supermarket chains*, Elsevier, Doetinchem.

De Koster, M.B.M., M. van de Vendel, M. de Brito (2001), Return handling: An exploratory study of nine retailers warehousing, working paper, Rotterdam School of Management, Erasmus University Rotterdam

Doherty, N.F., F. Ellis-Chadwick, C.A. Hart (1999), Cyber retailing in the UK: the potential of the internet as a retail channel, *International Journal of Retail & Distribution Management* 27(1), p.22-36.

Ferrara, C., A. Sarner, C. Claps (2000), Transforming retailing to 'E-tailing', The Gartner Group.

Grover, V., A.H. Segars (1999), Introduction to the special issues: Electronic commerce and market transformation, *International Journal of Electronic Commerce* 3(4), p.3-9.

Jedd, M. (2000), Fulfillment: A crucial e-business challenge, *Logistics Management and Distribution Report*, April, p. E25-E26.

Jin, L., D. Robey (1999), Explaining cybermediation: An organizational analysis of electronic retailing, *International Journal of Electronic Commerce*, 3(4), p.47-65.

Kaipia, R. (2000), A literature survey about current issues in business-to-business e-commerce. Working paper Helsinki University of Technology.

Kämäräinen, V., J. Småros, T. Jaakola, J. Holmström (2001), Cost-effectiveness in the e-grocery industry, *International Journal of Retail & Distribution Management* 29(1), 41-45.

Kämäräinen, V., J. Saranen, J. Holmström (2000), How goods receipt affects e-grocery efficiency, Working paper Helsinki University of Technology.

Kambil, A., P.F. Nunes, D.Wilson (1999), Transforming the marketspace with all-in-one markets, *International Journal of Electronic Commerce*, 3(4), p.11-28.

Kern, T., A. Aztouti, S. van de Velde (2000), The impact of internet technology on brokerage and integration in supply chains, Working paper, Rotterdam School of Management, Erasmus University.

Macht, J. (1996), Errand boy, *Inc* 18(16), 61-66.

McCullough, S.S. (1999), *Mastering commerce logistics*, Forrester Research, Cambridge.

Morganosky, M.A. (1997), Retailing and the internet: a perspective on the top 100 US retailers, *International Journal of Retail & Distribution Management* Vol. 25, Nr. 11, p.372-377.

Morganosky, M.A., B.J. Cude (2000), Consumer response to online grocery shopping, *International Journal of Retail & Distribution Management* 28(1), p.17-26.

Palmer, J.W. (2000), Electronic commerce in retailing: Convenience, costs, delivery and price across retail formats, *Information Technology and Management* Vol. 1, p. 25-43.

Pflaum, A., C. Kille, M. Wilhelm, G. Prockl (2000), Consumer direct - The last mile, Fraunhofer Anwendungszentrum Verkehrslogistik und Kommunikationstechnik.

Punakivi, M., J. Saranen (2000), Identifying the success factors in e-grocery home delivery, Working paper Helsinki University of Technology.

Punakivi, M., J. Holmström (2001), Extending e-commerce to the grocery business – The need for reengineering fleet management in distribution, *Logistik Management* 2(1).

Pyke, D.F., M.E. Johnson, P. Desmond (2001), E-fulfillment, it's harder than it looks, *Supply Chain Management Review*, Jan/Feb, 26-32.

Ranchhod, A. and C. Gurău (1999), Internet-enabled distribution strategies, *Journal of Information Technologies* 14, 333-346.

Sakar, M.B., B. Butler, C. Steinfield (1995), Intermediaries and cybermediaries: A continuing role for mediating players in the electronic marketplace, *Journal of Computer-Mediated Communication* 1(3), http://www.ascusc.org/jcmc/vol1/issue3/ sakar.html.

Småros, J., J. Holmström, V. Kämäräinen (2001), New service opportunities in the e-grocery business, to appear in *International Journal of Logistics Management*

Tanskanen, K. (2001), Logistical strategies for electronic grocery shopping, Working paper, Helsinki University of Technology.

The Industry Standard (2000), www.thestandard.com (mid 2000)

Van der Laan, J.W. (2000), The retail economics toolkit, http://www. retaileconomics. com/aboutpub1/htm, (mid 2000)

Van Goor, A., M.J. Ploos van Amstel, W. Ploos van Amstel (1996), *Fysieke distributie*, Stenfert Kroese, Houten

Xie, B. and X. Wang (2001), Case study: eGuo.com, Proceedings of the 12th annual conference of the Production and Operations Management Society.

Distribution Planning with Specific Delivery Time Restrictions for the Handling of Electronic Customer Orders in Food- / Non-Food Retail Trade

Joachim R. Daduna

University of Applied Business Administration Berlin, Badensche Straße 50 – 51, D – 10825 Berlin, Germany

Abstract. With an increasing spread of electronic shopping in the (food / non-food) retail trade new problems in distribution planning emerge. This especially refers to the development of efficient operations for the delivery of electronically executed customer orders to exist in this trade channel at the market for long-term. An essential condition is in this context the delivery on the order day (or on a determined day) within a certain time interval. If wished, the customer should be promised also an as-exact-as possible delivery time. To handle these problems, which represents a specific model of vehicle routing, solutions are required. That complies with the demanded conditions to the one and guarantee to the other, that the operative expenditures (vehicle and personnel demand) can be held as low as possible. For these problems a suitable procedure is introduced, that shows a higher degree of freedom in the combinatorial processes and enables to gain a more cost efficient routing.

1 Distribution in Electronic Shopping

The increasing spread of electronic shopping in retailing is followed a constantly rising demand for services in distribution logistics by (cf. e.g., Nachtmann (1999); Nilsson (1999) or Szász (1999)). This refers to different areas of business-to-consumer trade, as food / non-food retail trade, mail order sale as well as direct sale (s. e.g., Bretzke (1999); Daduna (2000b) and Daduna (2002)). On the basis of the specific logistic requirements, which are marked through mainly small parcel sizes and spatially strongly distributed customer locations, market potentials arise, that essentially has to be assigned to the area of parcel services (cf. e.g., Clausen (1998)). Examples are (established) suppliers as in the classical retail trade *Wal-Mart Stores, Inc.* (see www.walmart.com), in mail order sale as *Otto Versand* (see www.otto.de) or *Quelle AG* (see www.quelle.de) (cf. Bliemel / Theobald (1999) and Palombo (1999)) and also for new suppliers, as *Dell Computer Corporation* (see www.dell.com) (cf. Dörffeldt (1999)) and *Amazon.com*, Inc. (see www.amazon.com).

Another problem yields for the food / non-food retail trade, as the necessary services in distribution logistics can not (or only very restricted) be carried out

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

within the existing bounds of classical parcel services. The essential cause for this situation results from the fact, that the needed services are largely locally and (mainly) short-term oriented deliveries. This structure originates from a displace ment of logistical functions in the purview of the supplier, that is taken over within the traditional shopping normally through the customer. So in the frame-work of electronic shopping the supplier in this case gets the responsibility for order picking of the entered customer orders and for carrying out the customer supply. Independently from the (formal) question, whether these services are produced with business-own means or are acquired by accordingly configured outside work within the scope of *contract logistic solutions* (cf. e.g., Maltz / Ellram (1997); Daduna (2002)).

To be able to run the electronic shopping middle- and long-term successfully as a (normally additional) trade channel, sufficient logistical capabilities must be guaranteed as the customer always has the possibility to compare this (new) channel with the traditional shopping. An essential point is on this occasion an efficient vehicle routing, that guarantees on the one side a *timely-as-possible* delivery for the customer, which is, however, connected also with a *reasonable (financial) expenditure* on the other side.

For this specific problem, a suitable solution procedure is introduced. It derived from an existing *vehicle scheduling procedure* applied *in public mass transit planning* (cf. Daduna / Mojsilovic / Schütze (1993) or Daduna / Völker (1997)). In the following, the specific requirements in order delivery are described and are analyzed at first, and, going out from these structures, an approach for the vehicle routing with specific delivery time restrictions is introduced. Afterwards a procedure to solve such problems is presented and, on the basis of a (demonstration-)example, explained in detail.

2 Requirements for Vehicle Routing in the Food / Non-Food Retail Trade for Electronic Shopping

With the design of distribution structures for electronic shopping in the food / non-food-retail trade the requirement for conceivably very short-term deliveries is the most important point on the one hand, as well as on the other hand, however, also a high punctuality to meet partially quite narrow time windows or fixed delivery times. Essential characteristics are these both to produce the needed logistical services, that influences the acceptance of this trade channel in a significant range. As far as (customer-owned) box systems or an inclusion of pick-up-points does not occur (cf. Diller (1999); Braun / Primer / Stache (2000) or Daduna (2002)) but a direct supply of the customers (homedelivery) has to be performed, a customer-oriented and, as well, cost efficient vehicle routing is necessary.

The fundamental objective while dealing with customer requirements must be at first to fulfil consequently fixed delivery times, as well as, only small deviations (in form of unclearly defined delivery time windows) can be accepted. For in this

case underlying vehicle routing problems form, besides the *traffic network data* and the *truck fleet data* (as essentially static basic data), *order-relating customer data* the input for the planning processes. The necessary information about the individual orders includes the *customer location*, the *delivery time*, the *order weights / volumes*, the expected time for the *delivery handling* as well as conceivably *additional activities* with the customer (payment handling, returns, etc.).

For the supplier (or an authorized logistic provider), yields the number of needed vehicles and with it also, the number of personnel as crucial cost factors in the vehicle routing. However, uniform target systems cannot be assumed in planning processes on the side of the suppliers as conceivably differing conditions can be given. A possible approach shows as *priority objective the minimization of the vehicle number* needed to carry out a given order quantity, including the *reduction* of the *total mileage* as an objective of subsequent rank. Such a target hierarchy offers itself this type of planning problems as the *fixed cost share*, that an additional vehicle causes can normally not be compensated by cost cuttings within the (operative) mileage. If on the other hand a certain truck fleet, that exists for carrying out the delivery operations, should be used, also an even allocation of work to the vehicles can be the priority objective, with an as extensive-as-possible reduction of total mileage as an additional lower ranked objective likewise.

Vehicle routing processes represents (*combinatorial*) *optimization problems*, which can be solved by using different approaches (cf. e.g., Domschke (1997, pp. 207) or Fisher (1995)). Based completely on *deterministic data* such solution procedures normally can show considerable *restrictions* on the (*combinatorial*) *degree of freedom*. So in the case of fixed delivery times solutions can appear, which show despite of a conceivably calculated *optimality* (for the *formalized problem*) in the (*operational*) *realization* distinct cost inefficiencies. For this reason the question arises in which form an increase of the (combinatorial) degree of freedom can be gained in order to get more cost efficient solutions. Introducing (fixed) *time windows* (cf. e.g., Desrosiers / Dumas / Solomon / Soumis (1995)) for delivery operations with, e.g., an expansion of up to two or three hours, that would enable better results, is from customers' point of view an unattractive and in the end also an unacceptable solution. So it must be looked for another approach taking into consideration customer wants (with precise-as-possible determined delivery times) and the necessary logistical expenditures.

Therefore, a procedure has to be developed to calculate the best solution based on determined delivery times, but allowing deviations in a certain range. If cost cuttings are gained by reducing the number of needed vehicles. However, for the customer it must be possible to restrict these deviations by defining fixed (*individual*) time windows for the delivery time. In opposite to vehicle routing methods using *soft time windows* (cf. e.g., Taillard / Badeau / Gendreau / Guertain / Potvin (1997)), which include specific *penalty functions* to avoid (admissible) *time window violations*, the main objective is not to attain *compromise solutions* by minimizing simultaneously total milage cost and penalty cost, but the minimization of the *needed number of vehicles*.

3 Problem Description

The considered approach for the vehicle routing is based on a two-stage procedure applying a *schedule first - cluster second-strategy*. In the first *step delivery sequences* are formed, based on all customer orders for a fixed delivery period (cf, e.g. the stated delivery conditions by the Reichelt AG, see www.reichelt-ag.de). The result is a set of with the *optimal number* of sequences, including all served customer locations within the framework of the given data structures, especially incorporating the connections between the different customer locations. In a second step, these sequences are assigned to the available *delivery facilities* (among others *regional warehouses* or (bigger) *shopping facilities*), in which the customer orders become picked. If only one delivery facility is given (*single-depot problems*), the sequences are completed to *routes* by adding *pull-out trips* (from the delivery facility to the first customer) as well as *pull-in trips* (of the last customer back to the delivery facility). If there are, on the other hand several facilities (*multi-depot problems*), to which the sequences must be assigned under consideration of spatially and / or capacity-related aspects to the different delivery facilities, the sum of the length of all needed pull-out and pull-in trips should be held as low as possible.

The input data must be subdivided, as mentioned above, in (static) basic data and actual order-related (customer-)data, while these data are provided for the planning process from different sources. Furthermore, the data structure depends on specific spatial conditions, so other aspects have to be respected, e.g., whether it is an inner-urban, a suburban or a rural service area. In principle it is to be assumed the following data demand:

- Basic data
  - (Road)traffic network (Route length, running times, time of day- and / or directional-dependent influences on running times, turning restrictions, etc.)
  - Truck fleet data (Number of vehicles, loading capacities, technical equipment, etc.)
  - Package data / container data
- Order-related (customer-)data
  - Customer location (Street, house number, possibilities for car stops, etc.)
  - Order volumes / order weight (or number of containers)
  - Delivery time
  - Service time (= Time-consuming factors for the delivery process, influenced (among others) by path length to the house entrance, floor number, availability of an elevator, additional activities, etc.)

Based on the traffic network-related data, the possible connections d_{ij} ($i,j = 1, .., m$) to the m customer locations (for a fixed delivery period) must be prepared. On principle time- or distance-related values can be used for these calculations, but however, in this case the basis should be formed of (derived) time-related data t_{ij} ($i,j = 1, .., m$), as the service time is defined as a time value. Besides, this approach

also offers the possibility to include *time of day*- and / or the *directional-dependent* influences on running times in detail, so that differentiated values $t_{ij}(l)$ $(i,j = 1, .., m, l = 1, .., q(i,j))$ are assigned to each relation (i,j) of the network, while the number of intervals $q(i,j)$ results from *the daily traffic density profile* for (i,j). As far as no sufficient specifications of the network data are given, an approach to vary running time structures can be gained in a certain scope with interval-related weights.

Respecting the availability of the delivery vehicles, different vehicle types V_k $(k = 1, .., p)$ and its number $N(V)_k$ $(k = 1, .., p)$ is important. Furthermore, the capacities of these vehicles, $\kappa(V)_k$ $(k = 1, .., p)$, have to be seen, as well as if necessary specific vehicle equipment. Taking into account the complexity of such planning problems and also the influences on an efficient fleet management, however, a vehicle type variety should be avoided. A largely unified equipment should be aimed at, if no specific transportation requirements (among others for frozen food), restrictions from traffic infrastructure and / or topographical conditions may cause a suitable differentiation.

The *service time* g_i $(i = 1, ..., m)$ must be calculated in the last step, based on the relevant factors for every order, while g_i is to be understood as the time, that evolves between the arrival and the departure of a delivery vehicle at the customer location i $(i = 1, .., m)$. The (wished) *delivery time* Z_i $(i = 1, .., m)$ is determined by customer i, as well as a (fixed) delivery time window $F_i = (f1_i, f2_i)$ $(i = 1, .., m)$ with $f1_i$ as beginning and $f2_i$ as end. With the time windows in this case customer-individual restrictions are established to limit (possible) deviations of the delivery time, which are allowed within the solution procedure in order to gain suitable savings with the distribution cost. Moreover, the *demand quantities* b_i $(i = 1, .., m)$ must be included, and also, having a look on the estimation of the accruing transportation volumes, the question of the package and / or container concepts must be taken into account.

As far as it should be necessary in individual application cases, additional restrictions have to be taken into account. This refers, e.g., to the inclusion of an explicit assignment of vehicle types for certain delivery if frozen food is part of the orders. For such situations a vehicle type-pendent separation in two routing problems, which has to be solved independently, must not necessarily provide an appropriate solution, as the restrictions frequently refer only to some individual orders, while there are no corresponding restrictions for the others. This means, that an one-sided interchangeability is given in such cases, because cold-storage vehicles can serve other orders while this is not possible vice-versa for other vehicles. Moreover, the solution procedure can be influenced in different manners. On the one hand, it is possible to determine fixed combinations of some orders to closed route elements (*giant tasks*), aiming at a well-calculated pre-selection in the routing process. On the other hand, if bigger areas have to be served, a *spatial clustering* can be performed by an allocation of customers to different (decentralized) delivery facilities (= depots). With a largely fixed assignment of crews to depots, also a bigger knowledge about (local) street networks can be attained. That re-

duces the risk of not necessary losses of time, resulting from the search for customer locations.

4 Solution Procedure

Basic data for the vehicle routing process are at first the given delivery activities for a fixed time interval, defined by service times g_i, delivery time Z_i, delivery quantities b_i as well as the running times $t_{ij}(l)$ between all customer locations. The vehicle capacities are not taken into account at this stage as only a restricted number of customer can be served within the delivery intervals, which are normally very limited in time (see, among others, the example of the Reichelt AG at www.reichelt-ag.de). If operationally inadmissible solutions should appear in some cases through capacity violations, these solutions must be corrected by the responsible planning staff within the screening of the computer-aided calculated results making use of interactive adaptations in the respective routes.

At first a feasible start solution is calculated, making use of a (classical) linear assignment algorithm (cf. e.g., Carpaneto / Martello / Toth (1988); Domschke (1995), pp. 208), that contains only delivery sequences, which guarantees all predetermined delivery times. To solve this problem, a connection matrix $C = c_{ij}$ (i,j = 1, .., m) is to be built, that contains all admissible combinations of delivery activities. These combinations must fulfil the following condition:

$$Z_i + g_i + t_{ij}(l) \leq Z_j \tag{1}$$

Condition (1) is the basis to come up with a (formally) feasible solution, over whose quality no statement is possible, especially regarding to the *unproductive* times in delivery operations, that results, for instance, from an early arriving at customer j. For this reason an additional qualitative valuation of the admissible combinations is necessary, which could, as for instance, be fixed over the running time $t_{ij}(l)$ and a waiting time w_{ij} (j = 1, .., m), that can be defined as follows:

$$w_{ij} = Z_j - (Z_i + g_i + t_{ij}(l)) \tag{2}$$

As assignment problems are normally based on a *minimization approach*, unproductive times can be avoided with accordingly differentiated weights of the waiting times w_{ij}. Under consideration of the quantitative condition (1) and the qualitative points of view, the following values for the connection coefficients c_{ij} (i,j = 1, .., m) emerge:

$$c_{ij} = \begin{cases} \in [0;\hat{c}] & (i,j) \text{ is an admissible solution} \\ \infty & \text{otherwise} \end{cases} \tag{3}$$

with $\hat{c} << \infty$.

The valuation of all inadmissible (or undesirable) combinations with "∞" causes, that these appear at a minimization approach in the solution merely in the not avoidable extent, so that the number of the *admissible connections* in the calculated (optimal) solution is *maximized*. This means, a minimization of the fleet size required for the delivery of a defined order volume that can be attained making use of an (classical) assignment method.

Going out from this operationally feasible (and formally optimal) solution it is tried in an iteration procedure to reduce the number of the required vehicles step-by-step (cf. Daduna / Mojsilovic / Schütze (1993)), allowing graduated shifts of the delivery times Z_i. Based on the vehicle number r, which is already reached in the last iteration, the actual iteration is carried out with a modified connection matrix $\tilde{C}(R)$, while R defines, which number of vehicles (= number of delivery sequences) should be reached. The (redefined) coefficients of $\tilde{C}(R)$ look like follows:

$$\tilde{c}_{ij}(R) = \begin{cases} 0 & (i, j) \text{ is an admissible combination without deviations} \\ \in]0; \hat{c}] & (i, j) \text{is an acceptable combination with deviatians} \\ \infty & \text{otherwise} \end{cases} \tag{4}$$

The connection matrix must be adjusted with every new iteration in accordance with the performed changes. The combinatorial problem, that in these cases has to be solved in an iteration procedure, represents an *modified assignment problem*, that is supplemented through an additional restriction (8) to force a determined number of delivery sequences. It has the following formulation:

$$Minimize \sum_{i=1}^{m} \sum_{j=1}^{m} \tilde{c}_{ij} \, x_{ij} \tag{5}$$

subject to:

$$\sum_{i=1}^{m} x_{ij} = 1 \qquad j = 1, ..., m \tag{6}$$

$$\sum_{j=1}^{m} x_{ij} = 1 \qquad i = 1, ..., m \tag{7}$$

$$r \leq R \tag{8}$$

with m as number of customers (= orders to be served) and r as number of the *attained delivery sequences* for the *pre-determined* value R. If condition (8) is not fulfilled, the iteration procedure stops.

After finishing of this first step, the determined sequences are assigned to the delivery facilities, while the sequences through the pull-out trips and pull-in trips are completed to routes. With the existence of several delivery facilities, an assignment of the calculated sequences has to be done in a complementary step

making use of a (classical) *transportation algorithm* (cf. e.g., Domschke (1995), pp. 112). Included in this step is the route constructing process. The available vehicles at each facility represents in this model the "supply" and the delivery sequences the "demand", while the "transportation cost" emerge from the sum of pull-out and pull-in trip to operate a sequence starting from / ending in a defined delivery facility.

The course of the solution procedure looks, sketched shortly, as follows:

- **Step 1:**
  Preparation of the data required for the optimization procedure.
- **Step 2:**
  Generating at first a feasible start solution with the start value $R = m$. The result, the number of the determined delivery sequences without deviations of the given delivery times, yields the value r.
- **Step 3:**
  The execution of the iterations begins setting automatically the start value $R = r\text{-}1$, i.e. R is the number of the sequences that should be reached at least in the respective iteration with help of the modified assignment approach, and preparing the connection matrix $\tilde{C}(R)$. If in the optimization procedure this value is gained (or conceivably with $r < R$ remain under the given target value), the next iteration starts. If this is not the case, step 4 follows.
- **Step 4:**
  Construction of the routes outgoing from the determined sequences (with only one delivery facility) or the assignment of the sequences to the delivery facilities and subsequent forming of the routes.

If the vehicle routing is carried out for several delivery facilities simultaneously, in practical application conceivably a strong pressure of time evolves to make the results available, as the routing process and, if necessary, the subsequent assignment of routes to the facilities is also interconnected with an allocation of the orders to the different facilities. At this place, a sufficient efficiency of the used planning tool has to be guaranteed, so that the time period between the (different daily) deadlines for ordering and the beginning of the corresponding delivery periods, that is mainly required for (administrative) order processing and order picking, has not to be extended.

5 Demonstration Example

The following introduced (small) demonstration example should clarify that no great changes of the delivery times show an obvious effect on the planning results. The example is based on altogether eight customers, who (going out from the customer wishes) within a period from 65 minutes should be supplied. Only one ve-

hicle type is available, which shows, however, a sufficient capacity, and it is able to supply if necessary also all customers on one single tour. The customer order data (delivery times and service times) that are relevant for the planning process are given in Table 1 (time windows are not included in this demonstration example), where as Table 2 shows the running times t_{ij} of the relevant connection trips (cf. for this condition (1) in section 4) between two customer locations (without consideration of time of day- and / or directional-dependent influences on the running times).

Sorting the delivery times in a non decreasing order, there are no allowable values in the low triangle matrix, while at the same time the upper triangle matrix normally shows a low density.

Tab. 1: Relevant customer data

Customer	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Delivery time	0	4	14	17	30	41	53	60
Service time	8	8	9	7	5	10	11	5

Tab. 2: Running times t_{ij} on the relvant relations (i,j) ($M \cong \infty$)

(i,j)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1		M	8	12	16	25	8	17
2			10	4	11	14	19	7
3				M	16	18	23	14
4					7	21	18	21
5						8	15	22
6							11	2
7								M
8								

In **Step 1** the cost coefficients c_{ij} for all admissible connections are calculated, based on the following formulation including running time and waiting time, which must be seen as the most important qualitative criteria to rate a connection (i,j):

$$c_{ij} = \left\lceil \frac{(t_{ij})^2 + (w_{ij})^2}{100} \right\rceil^2 \tag{9}$$

The resulting cost matrix C is shown in Table 3.

Tab. 3: Cost matrix C (M $\cong \infty$)

(i,j)	1	2	3	4	5	6	7	8
1		M	M	M	9	49	225	256
2			M	1	4	25	81	324
3				M	M	16	36	64
4					M	M	25	49
5						M	9	25
6							M	1
7								M
8								

Based on these data, **Step 2** is carried out to determine a (first feasible) start solution considering (fixed) delivery times. Figure 1 shows the gained result with a (formal) optimal schedule, making the use of three vehicles a necessary fact. If one analyzes this first solution, it becomes obvious, that this result is not very favourable.

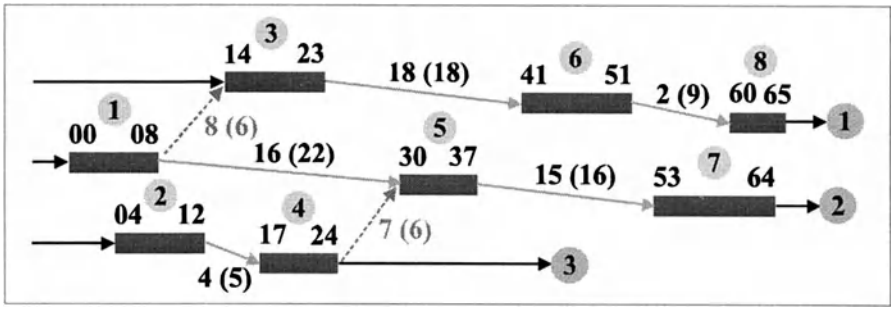


Fig.1: Start solution

Applying **Step 3**, it is aimed to reduce the number of delivery sequences (= needed vehicles), in this case from 3 (= r) to 2 (= R). At the beginning of the first iteration shifts, which allows only one minute in order to extend the combinatorial space. But it turned out, that these admissible deviations do not lead to an improvement. Therefore, shifts up to two minutes become included, so that in accordance with condition (4) (see section 4) the following cost matrix $\tilde{C}(R)$ is attained (see Tab. 4).

Alterations concerning the delivery times of customers 3, 4 and 6 are carried out, while customer 4 is shifted backward by one minute and customer 3 and 6 are shifted forward by two minutes, new combinatorial possibilities evolve. These lead to the solution presented in Figure 2, showing a schedule with only two delivery sequences. Because it is not possible to gain further reductions, **Step 4** is carried out. The result shows, that already marginal alterations can lead to distinct

improvements in the number of needed vehicles, that are expressed in a sufficient reduction of distribution cost.

Tab. 4: Cost matrix $\tilde{C}(R)$ $(M \cong \infty)$

(i,j)	1	2	3	4	5	6	7	8
1		M	10	M	0	0	0	0
2			M	0	0	0	0	0
3				M	M	10	0	0
4					1	M	0	0
5						M	0	0
6							M	0
7								M
8								

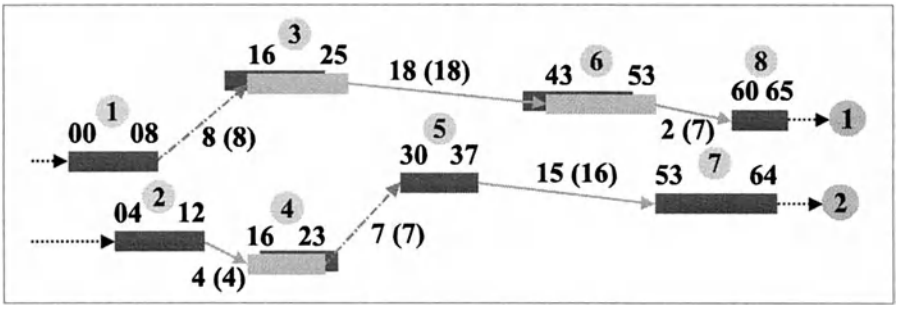


Fig. 2: Improved soution

As far as the deviations of the (wished) delivery times, as presented at the example, are varying in the range of a few minutes, these can perceived as "normal" fuzziness within distribution operations, and therefore these deviations are partially somehow not realized from customers. From which degree of deviation it becomes a must to inform a customer explicitly, however, should be reflected upon, in order not to generate a feeling of waiting (and with it also of unpunctuality). As the electronic (network) addresses of the customers are also available, resulting from the electronic order processes normally, it offers the opportunity of informing the customer about the delivery time planned for him as early as possible (i.e. after finishing the vehicle routing procedure).

6 Outlook

These descriptions show that the presented form of the vehicle routing procedure leads to suitable solutions for the increasing requirements in distribution planning in the area of the food - / non-food retail trade with electronically carried out orders. An essential point is in this case the comparison of *service quality* (from customers' point of view), represented through an as exact as possible adherence of delivery times, and *cost orientation* in delivery operations (from suppliers' point of view), represented on the basis of the willingness, as far as it is proved as cost efficient to accept also shifts for the delivery time within a certain range. Applications in the last years of this solution concept to vehicle scheduling problems in public transit led to significant cost savings in different mass transit companies (cf. e.g., Daduna / Mojsilovic / Schütze (1993) or Daduna / Völker (1997)), that are to be expected also concerning the presented distribution problems.

As a next step, it is intended to verify the propositions connected with this solution concept on the basis appropriate examples (with real world data) from the retail's area. In this case also a comparison with schedules, that were generated making use of already existing software tools, would be of interest as well as with results achieved manually. If it succeeds to prove the expected efficiency of this solution approach in detail, a tool for computer-aided routing has to be developed in another step, that is suitable also for end-user. The algorithmic basis for such tool exists (see Daduna / Völker (1997)), so that the essential task will be in the area of designing an efficient data management system and an user-friendly interface. Another aspect can be in this case the question of an attractive customer information concept, to the one respecting information about the planned delivery time, and to the other also, concerning to the current status of a delivery, as far as the delivery operations are monitored and steered with help of a suitable fleet management tool (cf. e.g., Lasch / Janker (1999) or Daduna (2000a)).

References

Bliemel, F. / Theobald, A. (1999): Der Einsatz des Electronic Commerce im Versandhandel. in: Tomczak, T. / Belz, C. / Schögel, M. / Birkhofer, B. (eds.): *Alternative Vertriebswege.* (Thexis) St.Gallen, pp. 322 - 339

Braun, A. / Priemer, J. / Stache, U. (2000): Vertrieb via Internet: Anforderungen und Auswirkungen. in: *Distribution* 31(6), pp. 8 - 11

Bretzke, W.-R. (1999): Smart Shopping im Internet - Industrie und Handel im Zeitalter von Electronic Commerce. in: Kopfer, H. / Bierwirth, C. (eds.): *Logistik Management.* (Springer) Heidelberg et al., pp. 221 - 243

Carpaneto, G. / Martello, S. / Toth, P. (1988): Algorithms and codes for the assignment problem. in: Hammer, P.L. / Simeone, P. / Toth, P. / Gallo, G. / Maffioli, F. / Pallottino, S. (eds.): *Fortran codes for network optimization.* (Baltzer) Basel, pp. 193 - 223

Clausen, U. (1998): Kurier-, Expreß- und Paketdienste. in: Buchholz, J. / Clausen, U. / Vastag, A. (eds.): *Handbuch der Verkehrslogistik.* (Springer) Berlin et al., pp. 64 - 70

Daduna, J.R. (2000a): Logistische Strukturen und Prozesse. in: Ammann, P. / Daduna, J.R. / Schmid, G. / Winkelmann, P.: *Distributions- und Verkaufspolitik.* (Fortis) Köln, pp. 281 - 310

Daduna, J.R. (2000b): Distribution bei Business-to-Consumer-Geschäften im Electronic Commerce. in: Inderfurth, K. / Schenk, M. / Ziems, D. (Hrsg.): *Logistik 2000plus - Herausforderungen, Trends, Konzepte* (Proceedings 6. Magdeburger Logistiktagung). pp. 122 - 133

Daduna, J.R. (2002): Distributionsgestaltung im Electronic Shopping. in: Conrady, R. / Jaspersen, T. / Pepels, W. (eds.): *Marketing Online Instrumente.* (Luchterhand) Neuwied pp. 241 - 261

Daduna, J.R. / Mojsilovic, M. / Schütze, P. (1993): Practical experiences using an interactive optimization procedure for vehicle scheduling. in: Du, D.-Z. / Pardalos, P.M. (eds.): *Network optimization problems - Algorithms, complexity and applications.* (World Scientific) Singapore et al., pp. 37 - 52

Daduna, J.R. / Völker, M. (1997): Vehicle scheduling with not exactly specified departure times. in: Preprints 7th International Workshop on Computer-aided Scheduling of Public Transport (Center of Transportation Studies Massachusetts Institute of Technology) Cambridge Mass., pp. 2 –12

Desrosiers, J. / Dumas, Y. / Solomon, M.M. / Soumis, F. (1995): Time constrained routing and scheduling. in: Ball, M.O. / Magnanti, T.L. / Momna, C.L. / Nemhauser, G.L. (eds.): *Network routing*, (North-Holland) Amsterdam et al., 1995, pp. 35 - 139

Diller, H. (1999): Die Akzeptanz von Zustell-Services für Konsumenten im Lebensmittelhandel. in: *logistik management* 1, pp. 198 - 207

Dörffeldt, T. (1999): Erfolgreicher PC-Direktvertrieb im Internet - Das Beispiel Dell Computer. in: Hermanns, A. / Sauter, M. (eds.): *Management Handbuch Electronic Commerce.* (Vahlen) München, pp. 405 - 409

Domschke, W. (1995): *Logistik*: *Transport.* 4. völlig überbearb. u. wesentl. erw. Aufl. (Oldenbourg) München, Wien

Domschke, W. (1997): *Logistik*: *Rundreisen und Touren.* 4. völlig neu bearb. Aufl. (Oldenbourg) München, Wien

Fisher, M. (1995): Vehicle routing. in: Ball, M.O. / Magnanti, T.L. / Momna, C.L. / Nemhauser, G.L. (eds.): *Network routing*, (North-Holland) Amsterdam et al., 1995, pp.1 – 33

Lasch, R. / Janker, C.G. (1999): Wirtschaftliche Aspekte der Telematik im Straßengüterverkehr. In: *logistik management* 1, pp. 209 -220

Maltz, A.B. / Ellram, L.M. (1997): Total cost of relationship: An analytical framework for logistics outsourcing decision. in: *Journal of Business Logistics* 18, pp. 45 - 65

Nachtmann, M. (1999): Elektronischer Geschäftsverkehr im Einzelhandel. in: Gora, W. / Mann, E. (eds.): *Handbuch Electronic Commerce.* (Springer) Berlin et al., pp. 311 - 326

Nilsson, R (1999): Einsatz und Potential von Internet-Shopping-Malls - Das Beispiel myworld. in: Hermanns, A. / Sauter, M. (eds.): *Management Handbuch Electronic Commerce.* (Vahlen) München, pp. 371 - 385

Palombo, P. (1999): Chancen und Risiken des elektronischen Versandhandels - Das Beispiel Quelle. in: Hermanns, A. / Sauter, M. (eds.): *Management Handbuch Electronic Commerce.* (Vahlen) München, pp. 361 - 369

Szász, T. (1999): Consumer Direct - Food-Lieferdienste auf dem Weg zu einem neuen Handelskanal. in: Tomczak, T. / Belz, C. / Schögel, M. / Birkhofer, B. (eds.): *Alternative Vetriebswege.* (Thexis) St.Gallen, pp. 360 - 387

Taillard, É. / Badeau, P. / Gendreau, M. / Guertain, F. / Potvin, J.-Y. (1997): A tabu search heuristic for the vehicle routing problem with soft time windows. in: *Transportation Science* 31, pp. 170 - 186

Chapter 4

Warehouse Location and Network Planning

An Analysis of a Combinatorial Auction

Mette Bjørndal and Kurt Jørnsten

Department of Finance and Management Science, Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 BERGEN, Norway

Abstract. Our objective is to find prices on individual items in a combinatorial auction that support the optimal allocation of bundles of items, i.e. the solution to the winner determination problem of the combinatorial auction. The item-prices should price the winning bundles according to the corresponding winning bids, whereas the bundles that do not belong to the winning set should have strictly positive reduced cost. I.e. the bid on a non-winning bundle is strictly less than the sum of prices of the individual items that belong to the bundle, thus providing information to the bidders why they are not in the winning set. Since the winner determination problem is an integer program, in general we cannot find a linear price-structure with these characteristics. However, in this article we make use of sensitivity analysis and duality in linear programming to obtain this kind of price-information. Finally, it is indicated how such prices can be used to enhance economic efficiency in an iterative market design. Throughout, the ideas are illustrated by means of numerical examples.

1 Introduction

In some auctions/markets, a participant's valuation of an object depends significantly on which other objects the participant acquires. Objects can be substitutes or complements, and the valuation of a particular bundle of items may not be equal to the sum of the valuations of the individual items, i.e. valuations are not additive. This may be represented by letting bidders of the auction have preferences not just for particular items, but for sets or bundles of items as well. In this setting, economic efficiency is increased by allowing bidders to bid on combinations of objects, which is exactly what a combinatorial auction does.

A recent survey of combinatorial auctions is provided by de Vries and Vohra (2000), also an excellent overview is given by Parkes (2001). In the literature, there are a number of examples of combinatorial auctions, ranging from the allocation of rights to radio frequencies (FCC (1994)), auctions for airport time slots (Rassenti et al. (1982)), railroad segments (Brewer (1999)) and delivery routes (Caplice (1996)). Bundle pricing (Hanson and Martin (1990)) and the effects of discounts on vendor selection (Moore et al. (1991)) can also be analyzed within this framework.

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

2 The Winner Determination Problem

Given a set of bids for subsets of objects, selecting the winning set of bids is denoted "the winner determination problem". This problem can be formulated as an integer programming problem. Let N be the set of bidders, M the set of m distinct objects, and S a subset of M. Agent j's ($j \in N$) bid for bundle S is denoted by $b^j(S)$, and we let

$$b(S) = \max_{j \in N} b^j(S)$$

The binary variable x_S is equal to 1 if the highest bid on S is accepted, and 0 otherwise. The winner determination problem can then be formulated as

$$\begin{aligned} \text{(IP1)} \qquad \max \quad & \sum_{S \subset M} b(S) \cdot x_S \\ \text{s.t.} \quad & \sum_{S \ni i} x_S \le 1 \qquad \forall i \in M \\ & x_S = 0/1 \qquad \forall S \subset M \end{aligned}$$

In some formulations of the winner determination problem, there is also a restriction stipulating that every agent/bidder can only receive at most one bundle. If we let binary variable $x^j(S)$ be equal to 1 if agent j receives bundle S and 0 otherwise, the corresponding formulation of the winner determination problem is the following

$$\begin{aligned} \text{(IP2)} \qquad \max \quad & \sum_{S \subset M} \sum_{j \in N} b^j(S) \cdot x^j(S) \\ \text{s.t.} \quad & \sum_{S \ni i} \sum_{j \in N} x^j(S) \le 1 \qquad \forall i \in M \\ & \sum_{S \subset M} x^j(S) \le 1 \qquad \forall j \in N \\ & x^j(S) = 0/1 \qquad \forall j \in N, S \subset M \end{aligned}$$

In both formulations the objective function maximizes the "revenue", i.e. the value of the bids, whereas the first set of constraints requires that no object can be assigned to more than one bidder. The second set of restrictions in (IP2) guarantees that no agent is assigned more than one bundle. An alternative interpretation of the maximization problems is the following: If bidders submit their true values, i.e. bid their reservation prices on different bundles, implying that $b^j(S) = v^j(S)$, for all $j \in N$ and $S \subset M$, the solution to the winner determination problem represents the efficient allocation of indivisible objects in an exchange economy.

Formulation (IP1) is valid for the winner determination problem in case of superadditive bids, i.e. if

$$b^j(A) + b^j(B) \le b^j(A \cup B) \qquad \forall j \in N,\ A, B \subset M \text{ and } A \cap B = \emptyset$$

In case of substitutes, as shown in de Vries and Vohra (2000) dummy goods can be introduced to make the formulation valid, or the more general formulation

(IP2) can be used. In any case, the formulation of the winner determination problem is an instance of the set packing problem (SPP). The linear programming relaxation of the SPP produces integer solutions in a number of cases (ref. de Vries and Vohra (2000)). We will however, focus on instances where the LP-relaxation gives fractional solutions. In general, the SPP belongs to the class of NP-hard problems, and is closely related to set partitioning and set covering problems.

Since in general the LP-relaxation produces fractional solutions, it is obvious that a set of market clearing linear prices need not exist for a combinatorial auction. This has led to the development of a number of stronger formulations of the winner determination problem. Bikchandani and Ostroy (1999, 2000) have presented two stronger formulations of (IP2). The first one is obtained by introducing artificial variables $y(k)$, and replacing the set of constraints requiring that each agent can obtain at most one bundle with the alternative set of constraints

$$\text{(LP1)} \qquad \sum_{j \in N} x^j(S) \le \sum_{k \ni S} y(k) \qquad \forall S \subset M$$

$$\sum_{k \in K} y(k) \le 1$$

where K is the set of all possible partitions, or "bundlings", of the items in M, and $k \ni S$ indicates that bundle S is represented in partition k.

This lead to a stronger problem formulation, in the sense that some of the fractional solutions that are feasible in the LP-relaxation of the weaker formulation, are cut off. However, the linear programming relaxation of this problem can still produce fractional optimal solutions. Another problem with this formulation is that the value of the dual is the sum of the maximal utility to each agent with bundle prices $p(S)$, plus the auctioneers maximal revenue. The use of *bundle* prices makes the price mechanism more complicated, and we are in this paper looking for a simpler evaluation scheme.

In the strongest formulation of the winner determination problem, the disaggregation goes even further, by replacing the constraints discussed above, with the constraints

$$\text{(LP2)} \qquad x^j(S) \le \sum_{k \ni [j,S]} y(k) \qquad \forall j \in N, S \subset M$$

$$\sum_{k \in K'} y(k) \le 1$$

where K' is the set of all possible *agent*-partitions, i.e. all possible combinations of "bundlings" and their allocation to different agents, and $k \ni [j,S]$ indicates that agent-partition k contains bundle S, which is allocated to agent j.

This formulation possesses the integrality property and hence, the linear programming relaxation is integer. However, the value of the dual becomes even more complicated since it is the maximal utility to each agent with bundle prices $p^j(S)$ plus the auctioneers maximal revenue over all feasible allocations at the prices. Note that the bundle prices $p^j(S)$ are non-anonymous or discriminatory bundle prices, i.e. every agent may face a unique vector of bundle-prices, making the evaluation even more complicated.

The problem as we see it with the two stronger formulations and their duals, is that they lead to non-linear price structures, with prices of objects and prices for bundles, that make it difficult to use them in a market mechanism design. In this paper we will present an alternative set of non-linear prices, that can be used to evaluate bids and give information back to the bidders/agents that can be used easily to determine the prospects of a bid-increase, or explain easily why a particular bid did not win.

3 Suggested Method

Consequently, the objective of this paper is not to focus on solution methods for the winner determination problem, but rather to find prices on individual items that support the optimal allocation of bundles of items. By "support", we mean that the prices on the individual objects should price the winning bundles according to the winner bids, whereas the bundles that do not belong to the winning set, should have strictly positive reduced cost, i.e.

bid on non-winning bundle	<	Σ prices of individual objects that belong to bundle

Prices with these characteristics will provide information to the bidders why they are not in the winning set, and this information may be used in a specific market design. Since the winner determination problem is an integer problem, in general, we will have to consider non-linear price structures.

It is only possible to find a single price-vector that excludes all non-winning bids if 1) the LP-relaxation of the winner determination problem has an integer solution, and 2) the LP-relaxation has a unique dual solution such that every non-winning bundle has reduced cost (RC) > 0. As will be illustrated in the next sections by means of a simple example, it seems to be difficult to find a unique price-vector with the characteristics searched for. Therefore, in this article, we suggest making use of sensitivity analysis to obtain this kind of price-information. The results of the sensitivity analysis are used to reduce the size of the (primal) winner determination problem and obtain a degenerate dual of the linear programming relaxation of the reduced primal. This generates a convex set of price-vectors such that RC > 0 for at least one price-vector for all non-winning bids. When reducing the primal, we search for a minimal reduction of the problem, in order to retain as much information as possible in the problem.

In the illustrative example we will use the first formulation of the winner determination problem presented (IP1), i.e. without the restriction that an agent can receive at most one bundle. However, we give other examples using the alternative formulation (IP2), taken from Parkes (2001), and illustrate the applicability of our proposed non-linear pricing scheme for this formulation as well.

4 Examples

In the first example, we assume that the following 9 bids have been handed in for different combinations of 7 objects, A-G:

Bid	17	10	10	9	20	12	4	15	26
Object									
A	1	1	1		1		1	1	
B		1	1					1	
C		1			1	1		1	1
D				1					
E	1		1		1	1			1
F		1	1				1		1
G	1			1	1			1	1

The interpretation of the table is as follows: the bid of 17 includes objects A, E and G, the next bid of 10 is on objects A, B, C and F, etc.

The winner determination problem of the combinatorial auction can be formulated as the following set packing problem, which is an instance of (IP1):

$$
\begin{array}{lccccccccccc}
\max & 17x_1 + & 10x_2 + & 10x_3 + & 9x_4 + & 20x_5 + & 12x_6 + & 4x_7 + & 15x_8 + & 26x_9 & & \\
\text{s.t.} & x_1 + & x_2 + & x_3 + & & x_5 + & & x_7 + & x_8 & & & \leq 1 \\
& & x_2 + & x_3 + & & & & & x_8 & & & \leq 1 \\
& & x_2 + & & & x_5 + & x_6 + & & x_8 + & x_9 & & \leq 1 \\
& & & & x_4 & & & & & & & \leq 1 \\
& x_1 + & & x_3 + & & x_5 + & x_6 + & & & x_9 & & \leq 1 \\
& & x_2 + & x_3 + & & & & x_7 + & & x_9 & & \leq 1 \\
& x_1 + & & & x_4 + & x_5 + & & & x_8 + & x_9 & & \leq 1 \\
& x_i & \text{binary} & & & & & & & & &
\end{array}
$$

The optimal integer solution has a value of 26, $x_9 = 1$ and all other variables are zero. In the following, we will consider various potential price-structures, based on 1) the LP-relaxation and 2) using sensitivity analysis together with linear programming. In the next section we will consider the use of IP marginal values.

1) LP-relaxation

If we relax the integer restrictions on the variables and solve the corresponding linear program, we obtain a fractional solution with value 26.5, where $x_1 = x_2 = x_9 = 0.5$ and all other variables are equal to zero. The shadow prices for the seven constraints are given by the vector (0.5, 0, 6, 0, 7.5, 3.5, 9) implying reduced costs for the 9 bundles that have been bid on equal to (0, 0, 1.5, 0, 3, 1.5, 0, 0.5, 0). However, this dual solution is not very useful in combinatorial auction terms, since it produces reduced cost equal to zero for a number of inferior bids.

This is so for bids 1 and 2, that is part of the fractional solution, but it is also so for bids 4 and 7, which are inferior even in the LP-relaxation.

Note that there exist multiple dual solutions to the linear program. The alternative dual solutions are (0.5, 0, 5, 0, 7, 4.5, 9.5) and (0.5, 0, 6, 0, 6, 3.5, 10.5), with reduced costs for the nine bundles equal to (0, 0, 2, 0.5, 2, 0, 1, 0, 0) and (0, 0, 0, 1.5, 3, 0, 0, 2, 0), respectively. We notice that for all the alternative dual solutions, several inferior bids have reduced cost equal to zero, but not necessarily in all the alternative solutions. Only the inferior alternatives consisting of bids 1 and 2 do not get any indications of the inferiority of the value of their bids, which is reasonable since they are part of the fractional LP-solution.

2) Using Sensitivity Analysis and Linear Programming

Let us use the example as we try to derive a system of prices that gives valuable information to the bidders, by just solving a number of linear programming problems. One alternative that first comes to mind, and that can generate valuable information to the bidders, is to perform a sensitivity analysis of each bid in turn, based on the assumption that the other bidders do not change their bids.

The information we are looking for is, *by how much must bidder i rise his bid in order to be guaranteed to get his bid accepted, given that the other bidders do not change their bids?* One way to get that information is to solve the parametric linear program until the variable corresponding to bid i takes the value 1 in the linear program. Note that the corresponding solutions for the other variables need not be integer, hence this value is just an indication of the necessary rise for bid i.

In our example we get the following results:

Bidder	1	2	3	4	5	6	7	8	9
Original bid	17	10	10	9	20	12	4	15	26
New bid	36	25	23	11	30	14	6	36	27

It is noteworthy to see that the competing coalitions consisting of bidders 4, 6 and 7 and bidder 9 get low rises, whereas the stand-alone bidders get high rises. However, this information is far from fair since bidder 3, by rising his bid to 18, while all other bidders keep their bids constant, would be in a winning combination together with bidder 4.

Question: Is there a way to generate a non-linear price-system by solving only linear programs?

What we are looking for is a price-system that yields reduced costs that are positive for all bids that are not present in the optimal solution to the winner determination problem. One way of finding such a system of prices is to reduce the winner determination problem by deleting bids such that the reduced winner determination problem, when solved as a linear program, yields the integer programming solution. That this can always be achieved is obvious, since we can reduce the winner determination problem to only include the winning bids. However, doing such a radical reduction will give us a price-system with very little information. Performing the maximal reduction in the example, with only the

winning bid left, gives alternative optimal dual price-vectors equal to (0,0,26,0,0,0,0), (0,0,0,0,26,0,0), (0,0,0,0,0,26,0) and (0,0,0,0,0,0,26). All deleted bids will have strictly positive reduced cost for at least one of the price-vectors. However, the price-structure indicates for the bidders that all of them need to rise their bids to 26, which is clearly a fraud.

In the following, we find a minimal reduction of the winner determination problem, that yields the result sought for, i.e. we delete as few as possible of the most probably losing bids, such that the reduced problem has an integer solution to the LP-relaxation. In the example, we achieve this by deleting bids 1, 2, 3 and 8. Solving the restricted LP-problem gives the winning bid, bid 9, and a set of 6 extreme dual solutions. The extreme dual solutions are given in the columns of the following table:

	Dual Solutions π					
Constraint	1	2	3	4	5	6
A	0	0	0	0	0	0
B	0	0	0	0	0	0
C	0	13	12	0	12	0
D	0	0	0	0	0	0
E	13	0	0	12	0	12
F	4	4	4	4	5	5
G	9	9	10	10	9	9

If we require that all bids should be able to tackle each of these prices and all the convex combinations of them, we can compute the maximal reduced costs for the 9 bids. This gives the following:

Bid	1	2	3	4	5	6	7	8	9
Reduced cost	5	7	7	1	2	1	1	7	0

As is evident from the numbers, the suggested price-structure prices out all non-winning bids, and the maximal reduced costs also provide a realistic indication of the necessary rise for each bidder, in order to be in the winning set.

In his Ph.D. thesis, Parkes (2001) presents a set of illustrative examples. One of the examples is as follows:

	Bundles						
	A	B	C	AB	BC	AC	ABC
Bidders							
Agent 1	60	50	50	200	100	110	250
Agent 2	50	60	50	110	200	100	255
Agent 3	50	50	75	100	125	200	250

The numbers in the table give the bids for the various bundles, from each agent or bidder. In this example it turns out that the linear programming relaxation of this instance of (IP2) has the optimal solution $x^1(AB) = x^2(BC) = x^3(AC) = 0.5$ with value 300, whereas the integer solution and hence optimal solution to the winner determination problem (IP2) is $x^1(AB) = x^3(C) = 1$, with value 275.

Applying our procedure to this problem, we get a minimal reduced formulation by deleting the bid from agent 2 on bundle BC and the bid from agent 3 on bundle AC. The corresponding dual problem has massive dual degeneracy, and hence there exist many alternative extreme dual prices. However, knowledge of only two of these prices is enough to price out all the non-winning bids. These prices are (140, 60, 75, 0, 0, 0) and (50, 130, 75, 20, 0, 0). The first 3 elements of the given price-vectors are prices for the three objects, whereas the 3 remaining elements are prices for the restriction that each agent can receive at most one object. This is an example where the LP-relaxation of the stronger formulation (LP1) possesses an integer solution, and the prices stated above constitute an alternative to the vector of *bundle*-prices (50, 60, 75, 190, 200, 200, 255).

A second example from Parkes (2001) is given by the following table:

	Bundles		
	A	B	AB
Bidders			
Agent 1	0	0	3
Agent 2	2	2	2

In this example, the optimal solution to (IP2) is given by $x^1(AB) = 1$, while the LP-relaxation gives $x^1(AB) = x^2(A) = x^2(B) = 0.5$. Only the strongest formulation (LP2) possesses the integrality property, and the non-anonymous / discriminatory vectors of bundle-prices resulting from the LP-relaxation of the strong formulation is $p^1 = (0, 0, 2.5)$ and $p^2 = (2, 2, 2)$. These prices should be compared to the two prices generated by our suggested procedure, which are (3, 0, 0, 0) and (0, 3, 0, 0), and which price out all the non-winning bids, without being non-anonymous.

5 IP Marginal Values

Acknowledging the discrete nature of the combinatorial auction problem, we will in this section, as a reference, investigate different IP marginal values, including

prices resulting from simple marginal calculations and prices resulting from making use of cutting planes and dual functions.

1) Exact IP Marginal Values

A way to generate price-information, although requiring a substantial amount of computation, is to calculate exact marginal values by solving a number of integer programming problems apart from the original integer program. These other integer programs to be solved, are generated by in turn adding one more unit of each object, or alternatively, deleting one object at a time, in each new problem to be solved. This may give an indication of the value of each object. In the first example described in section 4, this will give the following price information:

- (0, 0, 7, 0, 7, 5, 10), where each price is calculated by *deleting* the corresponding object (setting the right hand side equal to 0) and comparing the IP solution generated, with the original problem's IP solution. The reduced costs are given by (0, 2, 2, 1, 4, 2, 1, 2, 3), so this price-structure prices out x_2, but not x_1. Moreover, x_9 is not priced according to the winning bid, since the reduced cost for bundle 9 is different from 0.
- (0, 0, 5, 0, 5, 4, 9), where each price is calculated by *adding* a unit of the objects (setting the right hand side equal to 2) and comparing the IP solution value with the IP solution value of the original problem. In this case, neither x_1 nor x_2 are priced out, and the price-vector does not price bundle 9 to 26 either.

Consequently, these two price-vectors may indicate by how much each bidder might lower or rise his price, but do not give enough information to the bidders. Moreover, the procedure is computationally burdensome, as the number of IP solutions needed to generate the prices is 2 times *m*, the number of objects in the auction.

4) Cutting Planes and Dual Functions

We know that a price-system that works for a combinatorial auction must in general be non-linear. Such a non-linear price-system can be derived using a cutting plane or branch-and-bound technique when solving the winner determination problem. For our example we will use a cutting plane approach to generate a non-linear price-system. By adding constraints 1, 3 and 5, dividing by two, and rounding down, the following cutting plane is derived:

$$x_1 + x_2 + x_3 + x_5 + x_6 + x_8 + x_9 \leq 1$$

If we append this cutting plane to the winner determination problem and solve the new linear programming relaxation, we get the solution $x_9 = 1$ with value 26, shadow prices (0, 0, 0, 0, 0, 4, 9, 13), and reduced costs (5, 7, 7, 0, 2, 1, 0, 7, 0). I.e. bundles 1 and 2 are priced out, but not bundles 4 and 7.

There are however, multiple dual solutions, as shown in the table below:

Dual Solutions (π,μ)

Constraint	1	2	3	4	5	6	7
A	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0
C	0	5	0	4	5	0	4
D	0	0	0	0	0	0	0
E	0	7	0	7	6	0	6
F	4	4	4	4	4	5	5
G	9	9	10	10	10	9	9
Cut	13	1	12	1	1	12	2

Each of these dual solutions gives rise to a nonlinear price-function of the form

$$\mathbf{F}(\mathbf{a}_j) = \pi\mathbf{a}_j + \mu \cdot \lfloor (a_{1j}+a_{3j}+a_{5j})/2 \rfloor$$

where $\mathbf{a}_j$ is the coefficient vector of the jth bundle, π is the vector of shadow prices for the original constraints A-G, μ is the shadow price of the cut, a_{ij} is the coefficient in the ith row of bundle j, and $\lfloor u \rfloor$ represents the greatest integer less than or equal to u. However, none of the above dual non-linear price-functions alone will produce a positive reduced cost for all the unsuccessful bids. Nevertheless, a convex combination of the dual solutions yields such a price-system. Take for instance the convex combination consisting of 0.98 of dual solution 2 and 0.01 of dual solutions 3 and 6, giving prices (0, 0, 4.9, 0, 6.86, 4.01, 9.01, 1.22) and reduced costs (0.09, 0.13, 2.09, 0.01, 1.99, 0.98, 0.01, 0.13, 0).

This price-vector has the characteristics searched for, and this gives rise to the following questions:

Question 1: Does there always exist a set of cutting planes that produces prices such that all unsuccessful bids have RC > 0?

Note that from a pure integer programming point of view, this property is uninteresting, but when using integer programming in a combinatorial-auction-setting, this property is necessary in order to achieve decentrability.

Question 2: If such a set of cutting planes exists, how difficult is it to derive it, as compared to deriving only a set of cutting planes that yields the desired integer solution?

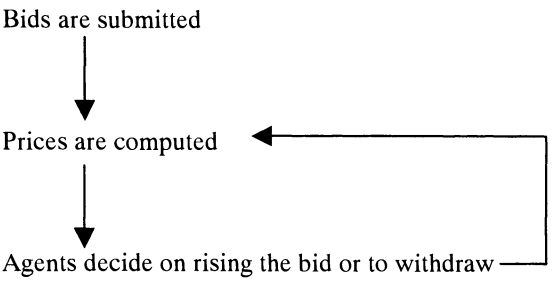
The two questions raised above are of theoretical interest. However, when it comes to implementing a combinatorial auction in practice, we need more easily accessible information to be distributed to the bidders. Thus, in this paper we have been interested in finding out whether such information can be derived, and if so, the next issue is in which form it should be distributed to the bidders, in order to create a market mechanism for a general combinatorial auction.

6 Outline of Market Design

Solving a combinatorial auction consists of two parts: 1) The Allocation Problem: "Who gets what?" and 2) The Pricing Problem: "How much do winners pay?" Bidders would naturally prefer to pay less than their reservation prices on their winning bids and this may induce strategic bidding. I.e. depending on how bids determine what a winner should pay, there may be incentives to bid less than the reservation price, $b^j(S) < v^j(S)$ for some j and S, in which case the auction is not incentive compatible.

In the literature, a number of incentive problems are discussed, including for instance "exposure problems" (bidders pay too much for individual items or drop out early in the bidding process in order to limit losses) and "threshold/free rider problems" (bidders may bid less than their reservation price with the aim of paying less, but risk losing the auction).

In order to increase economic efficiency, prices could be used in an iterative market design, where the agents can change their bids based on price-information, and hopefully, this process will produce a more efficient outcome. In the following, we indicate how the suggested price-structure can be used in a sequential combinatorial auction. The process could be the following:



When prices converge, in the sense that there is no more information to withdraw from the price-structures the auction could be closed by picking the previous integer solution as the "best" allocation.

If we consider again the first example of section 4, the following maximal reduced costs were computed based on the multiple dual solutions, for the 9 different bids:

Bid	1	2	3	4	5	6	7	8	9
Reduced cost	5	7	7	1	2	1	1	7	0

Assuming that the bidders rise their prices according to this information, i.e. no bidder withdraws because his reservation price has been exceeded, we get a revised winner determination problem. Only the objective function has changed compared to the old problem, as the bids have been increased by the maximal reduced cost. The new objective is

$$\max \quad 22x_1 + 17x_2 + 17x_3 + 10x_4 + 22x_5 + 13x_6 + 5x_7 + 22x_8 + 26x_9$$

Now the linear programming relaxation of the winner determination problem has objective function value 33.5 with $x_2 = x_3 = x_6 = 0.5$ and $x_4 = 1$, whereas the integer programming solution has value 28 with $x_4 = x_6 = x_7 = 1$.

Constructing once more the minimal reduced problem, requires that all bids except the winning bids, must be deleted from the problem in order to have an integer solution to the LP-relaxation of the reduced winner determination problem. Solving the reduced problem and calculating the corresponding price-structure yields

	Dual Solutions π							
Constraint	1	2	3	4	5	6	7	8
A	5	5	0	0	5	5	0	0
B	0	0	0	0	0	0	0	0
C	0	13	13	0	0	13	13	0
D	0	0	0	0	10	10	10	10
E	13	0	0	13	13	0	0	13
F	0	0	5	5	0	0	5	5
G	10	10	10	10	0	0	0	0

and

Bid	1	2	3	4	5	6	7	8	9
Reduced cost	6	1	1	0	6	0	0	6	2

Again, assuming that bidders rise their bids accordingly, the revised objective function is

$$\max \quad 28x_1 + 18x_2 + 18x_3 + 10x_4 + 28x_5 + 13x_6 + 5x_7 + 28x_8 + 28x_9$$

This winner determination problem has an LP-relaxation with value 37.5, whereas the integer solution has value 28. In fact, there exist alternative integer solutions, all with value 28, these are

1: $x_1 = 1$	3: $x_3 = x_4 = 1$	5: $x_5 = 1$	7: $x_9 = 1$
2: $x_2 = x_4 = 1$	4: $x_4 = x_6 = x_7 = 1$	6: $x_8 = 1$	

Since all variables appear in at least one of the alternative solutions, there does not exist a price-structure with the desired property. However, each bidder has now got the information on the relative value of their original bid.

If we instead use the information from the LP-relaxation of the original winner determination problem, i.e. base the reduced costs on the three dual solutions given in part 1) of section 4, our changed objective function will be

$$\max \quad 17x_1 + 10x_2 + 11.5x_3 + 10.5x_4 + 22x_5 + 13.5x_6 + 5x_7 + 17x_8 + 26x_9$$

which of course does not alter the bids on bundles 1 and 2. The corresponding solution to the linear programming relaxation of this perturbed winner determination

problem is integer valued, with value 29. Hence, the LP-prices make the procedure converge with a higher value paid to the auctioneer.

7 Conclusions and Suggestions for Future Research

In this paper, we have suggested a procedure for obtaining prices of individual items that support the optimal allocation of a combinatorial auction. This price-information is obtained by using sensitivity analysis and linear programming to delete bids from the allocation problem, thus creating a reduced problem that has an LP-relaxation that produces integer solutions and a degenerate dual. The multiple solutions to the dual problem generate several prices, and this price-structure has the properties sought for, namely that the winning bundles are priced according to the winning bids, whereas the bundles that do not belong to the winning set, have strictly positive reduced costs.

Such a price-structure provides information to the agents why they are not in the winning set and could be used in an iterative market design in order to enhance economic efficiency. The incentive-properties of an auction design like this is a topic for future research, as is also more extensive testing of the pricing-algorithm, and attempting to generalize the procedure to multi-unit auctions. However, the suggested method seems promising when it comes to providing information supporting decentralized decision making in a market setting.

References

Bikchandani, Sushil / Ostroy, Joseph M. (1999): The Package Assignment Model. Technical Report, Anderson Graduate School of Management and Department of Economics, U.C.L.A.

Bikchandani, Sushil / Ostroy, Joseph M. (2000): Ascending Price Vickrey Auctions. Technical Report, Anderson Graduate School of Management and Department of Economics, U.C.L.A.

Brewer, P. J. (1999): Decentralized Computation Procurement and Computational Robustness in a Smart Market. *Economic Theory*. 13;41-92.

Caplice, C. G. (1996): An Optimization Based Bidding Process: A New Framework for Shipper-Carrier Relationships. Thesis, Department of Civil and Environmental Engineering, School of Engineering, MIT.

de Vries, Sven / Vohra, Rakesh (2000): Combinatorial Auctions: A Survey. Technical Report, Kellogg Graduate School of Management, Northwestern University.

FCC Auction, Broadband Personal Communication Services, Bidder Information Package (1994): Washington DC.

Hanson, W. / Martin, R. K. (1990): Optimal Bundle Pricing. *Management Science*, 36;155-174.

Moore, E. W. / Warmke, J. M. / Gorban, L. R. (1991): The Indispensable Role of Management Science in Centralizing Freight Operations at Reynolds Metals Company. *Interfaces*, 21;107-129.

Parkes, David C. (2001): Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency. Doctoral Dissertation, Computer and Information Science, University of Pennsylvania.

Rassenti, S. J. / Smith, V. J. / Bulfin, R. L. (1982): A Combinatorial Auction Mechanism for Airport Time Slot Allocation. *Bell Journal of Economics*, 13;402-417.

The Practice of Distribution Network Planning: Coping with Shortcomings in Import Data Quality

Angela Bauer

Fraunhofer Anwendungszentrum für Verkehrslogistik und Kommunikationstechnik

Keywords. Application of Decision Support Systems, Data Quality

1 Introduction: Necessity and Challenge of Computer Based Logistical Network Planning

A rapidly changing, highly dynamic business environment, as well as the growing importance of worldwide cooperation of businesses in newly emerging supply chain networks are the reasons for the evolution of ever more complex logistical network structures. The ability to quickly adapt networks to new strategic challenges, as well as pressures for more operational efficiency demand for more frequent changes in network design and more efficiency in everyday network operations.

These trends explain why computer-based Decision Support Systems (DSS) are increasingly applied in the practice of logistics management, especially in consumer goods distribution and third/fourth party logistics service provider (3PL/4PL) industries. A critical property of computer based logistical network planning tools is the fact that large amounts of data must be handled.

The challenge of computer based logistical network planning can be described as trifold:

- First, the clients' problem has to be modelled adequately. Normally this means that network design and optimization have to meet very specific and highly complex client specifications.
- Second, pre-existing, standardized Decision Support Systems (DSS) have to be employed for reasons of cost effectiveness and time pressures that offer a variety of modelling options, but typically do not meet all the clients' demands.
- Third, to run the network DSS large amounts of mass „transaction" data must be handled which usually are characterized by a very poor quality.

Consequently the DSS-modellers' task is to cope with the poor quality of mass data and with limited modelling capabilites, while meeting clients' demands for high degrees of specificity, precision and quality of the solutions provided.

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

Most academic and professional work in the field of Decision Support Systems development in the past has focused on the algorithms applied – especially on their logical and mathematical qualities.

The practice of applying network DSS in „real" consulting projects commissioned by business world clients shows, however, that difficulties do primarily not ly in the algorithmic cores of the DSS. Instead they are caused by poor data quality and customer specifications that do not fit with features of standard DSS tools.

Therefore a significant, often greater challenge than algorithm development is the task to find and apply methods and heuristics to cope with import data quality during the modelling process and to find solutions how to intelligently adapt standard tools and standard approaches to highly client-specific problem definitions.

This paper discusses some of the experiences that the Nuremberg Fraunhofer DSS research group has made in this context. It is about modellers' coping strategies and tactics with problems prior to and outside of the issues of algorithm development and DSS-design.

The paper is organized in three steps: first, a typical network DSS tools that is used by the research group is briefly described in order to give the reader some background for the following arguments. The next part of the paper explicates the problems of poor data quality and modelling limitations using an example of the cosmetics industry, and how the research group was „coping" with these problems. In the following paragraph the same is done with an example of the foods industry. As a conclusion, some generalizations on coping strategies from the observations and arguments are drawn.

2 The Tool in Use: NC_{dis} as a Typical Modelling Application in Logistics of the Consumer Goods Industry

As a baseline for the following arguments, NC_{dis} (i.e. „network configuration for distribution problems) - a planning tool developed and used at the Nuremberg Fraunhofer DSS research group to generate decision support - will be explained to enable the reader to follow and assess the modelling examples presented below.

NC_{dis} is used to analyze, calculate and optimize complexe distribution networks (compare figure 1 and 2). Standard questions to be answered by using the tool are the following:

- Which number of facilities is best for a distribution network?
- Where should those facilities be located?
- What should be the storage or througput capacity of the facilities?
- How should customers be assigned to the facilities?
- What savings could be generated by running a cooperative distribution network with competing companies?
- How many deliveries per week and customer stand for an optimal structure?

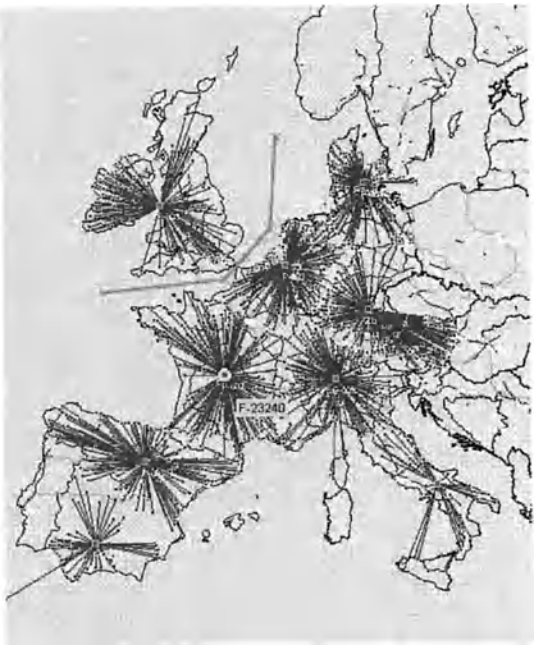


Fig. 1. NC $_{dis}$ – Customer Allocation to Facilities

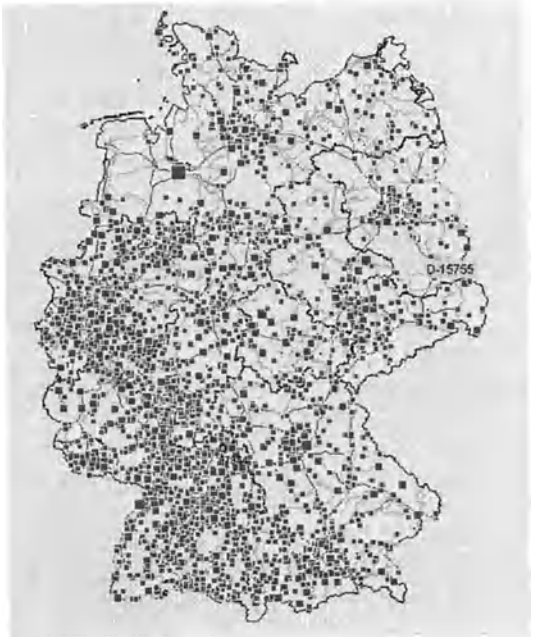


Fig. 2. NC $_{dis}$ – Proportion of Customers and weights

The tool, hence, supports practitioners in answering the classical location and allocation problems as well as in assessing the effects of the consolidation of independent networks into cooperative or merged networks.

The principal sequence of steps that has been found effective in the application of NC $_{dis}$ in consulting projects is the following:

- First: The underlying problem situation is identified jointly with the client, and a project plan is established.
- Second: A detailed analysis of the clients' order or shipment „transaction" data is done. The tool needs to be fed with a detailed shipment data base. Raw data provided by the client – usually from their order processing systems - is tested with respect to the completeness and plausibility of the data. Gaps and errors are corrected.
- Third: With the adjusted data base an initial calculation is started to establish a baseline to compare scenarios. The baseline should map the status quo as closely as possible to be able to estimate different alternative scenarios in any consequences. The gap between the status quo baseline and the client's perceptions is verified. Here it is important to gain information from as many client data sources as possible and to harmonize them to make sure that the quality of the status quo reconstruction is adequat.
- Fourth: After establishing the status quo baseline, several scenarios with modified network configurations – e.g. numbers or positions of locations, different distribution strategies, alternative costing schedules - are calculated and compared to each other.
- Fifth: By an interactive and incremental step by step process a solution that satisfies the client's and the modellers' optimization targets will be found.

The entire process is accompanied by intensive discussions and consultations with the practitioners to make sure to keep close to reality respectively to produce reasonable iplementable solutions.

3 First Example: Application Project in the Cosmetics Industry

3.1 Description of the Project – Objectives

The application case mentioned here took place in a large international cosmetics company with worldwide production sites and distribution channels. The objective of the project had been the optimization of the European distribution network. In the status quo situation the European distribution in the nine analyzed countries was organized in nine Distribution Centers (DC) – always one lying in one country. The main questions of interest for the company were the following:

- How many distribution centers were best for the companies European network with respect to cost and delivery time aspects?

- Which of the given DC locations should be kept and which of them should be given up or laid together with others?
- What about a „one Central European DC"-situation?
- Is the extension of capacities in several locations necessary?
- What about the tariff structures?

The primary objective had been a reduction in cost. The support of the above mentioned decisions with the DSS NC $_{dis}$ had been decided and data collecting and analyzing could start. After modelling the status quo situation as a baseline, several principal alternative network configurations had to be calculated and the corresponding consequences with respect to different criteria like cost, service, network flows and politics had to be shown up. As a result a strategic decision about the optimal amount and location of distribution centers and the optimal degree of centralization – decentralization had to be presented.

3.2 Coping with Shortcomings in Imort Data Quality

In this project several significant shortcomings in the quality of the order and shipment transaction data had to be overcome. Each of the nine European DC's had completely different data bases. The format as well as the quality of the data were not fully comparable. In some cases there was no information about the primary source of shipments . Moreover, in a lot of shipment data lines there was no weight assigned. Sometimes there were gross weights, sometimes net weights available with unknown tare factors. The problem was that there were different package systems in different countries meaning that the use of broad estimates to correct these inconsistencies was not possible. Each country had different data formats that had to be standardized. A logistics report that was based on a questionnaire regularly issued by the corporate logistics departement provided some important figures for each country. But there were many contrarinesses in this questionnaire. In one country data was not even available in any computerized format. One reason for these problems was company politics: The DC managers did not want to have anybody look into their operations in detail.

The first step on the way to optimize the European distribution system of the company, consequently, was an intensive effort in preparing the data base for analysis. Discussions were held with the practitioners. Visits of different locations as well as the creation of ad hoc records in warehousing systems, packaging, transportation strategies and data files for several countries were necessary.

After preparing a unified data file across all countries the data had to be completed and analyzed with respect to plausibility. First a separat article data base was created to be able to complete shipment data lines lacking article weights but containing article numbers. After this adjustment the same had to be done with a separat production file giving information about the source of articles with special article numbers. These article numbers were found again in the shipment data file where then the production site could be added on base of the production file. The so completed data file was discussed intensively with the practitioners and adjusted by examining the real processes at the factories.

The adjustment of the shipment data on base of the separat article data base and the production file helped to fill in lots of blanks in the shipment data available. But there were still missing lots of lines. For this reason a heuristic was generated that projected the information of the existing data base and evaluated new shipment data. The existing data base was analyzed with respect to weight clusters, plant production programs and customer profiles. All informations available in the company concerning this subject were gathered and put together. On base of these informations a „typical shipment strucuture" for the company was worked out and projected to generate artificial shipment data for the remaining data blanks.

To be able to fullfil this heuristic for the mass of several million sets of shipment data a special „mini" software tool was developed (compare figure 3). The available data for each country could be fed in and a realisticly projected and adjusted data base was put out.

Again, the revised data had to be verified in discussions with the practitioners and compared with the company's „official" logistics data set. This experience demonstrates how difficult and time consuming the task of preparing the data base can be, before an effective simulation and optimization procedure may start. Often even large, admired companies have problems in delivering adequate data quality. Quite frequently the problem is due to intentional „political" manoevring of the actors involved.

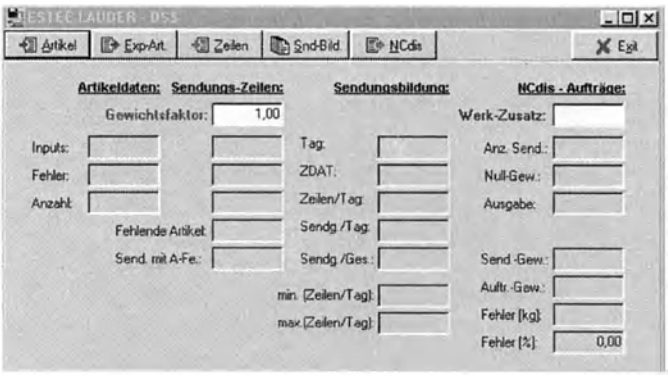


Fig. 3. Special mini software tool

3.3 Coping with Modelling Limitations

A second problem preliminary to the actual simulation and optimization concerns the „fitting" of the real world problem to the structures and capabilities of pre-existing modelling tools.

In the example described here, there were plants that had specific shipments for different countries. These shipments cannot be mixed because they differ from each other in labeling, language of package inscription, etc.

Another example could be that customers of special countries can only be served by certain distribution centers. This has above all political reasons but is

sometimes also connected with product policies. Consequently also the relations between the distribution centers underly specific political rules.

As a last point there exist lots of shipments from overseas that must enter the European distribution system at some point. The coping tactics a modeller could apply being confronted with the above mentioned specialities are the following:

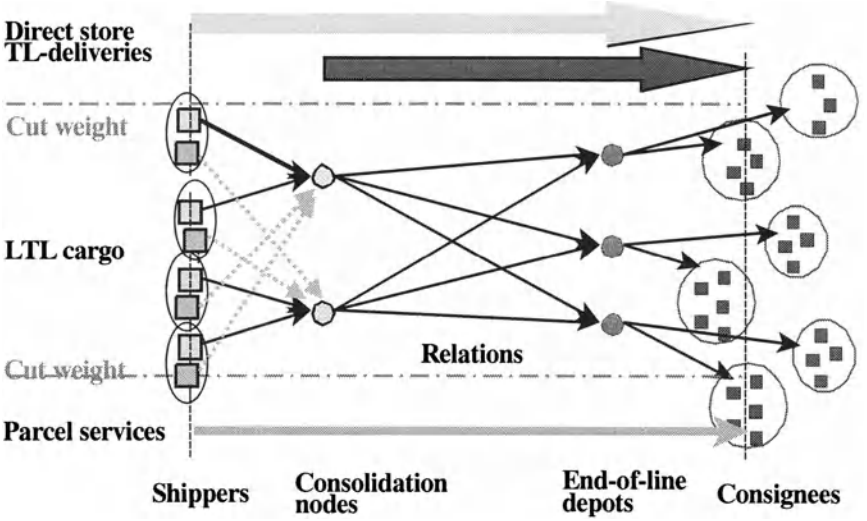


Fig. 4. „Splitting plants"

1. To be able to handle the specific plant alignments to certain countries or customers several plants or sources should be created at the same place (compare figure 4). This way of „splitting plants" means that one plant could be divided in several service areas. For example there exists one area that only delivers products to Great Britain or another that only serves German customers and so on. This splitting allows to cope with the tool limitation of being able to model only one simple plant or distribution center with no possibilities of differentiating countries or customers.
2. A tactic to cope with the problem of modelling shipments from overseas with the help of a European based DSS-tool is to introduce several virtual harbour points for example Rotterdam. For all shipments from overseas the source has to be changed in one of the virtual harbours when preparing the shipment data base. The way from the harbours to the distribution centers is calculated as pre-carriage. This is a method that handles the problem of a geographically limited model and at the same time is very close to the transport procedure in reality.
3. A third suggestion of modelling tactis is the definition of fixed relations between certain distribution centers. Especially if there is no information about tariff tables these relations should then be calculated with for example 5-to-assessory-rates for long haul transports as a fixed condition. This equals cost for full truck load for the consolidated „replenishment".

3.4 Results

The result that had been achieved by calculating the whole application project in intensive cooperation with the practitioners was the suggestion of several alternative network structures with savings up to 25%. This had met the formulated objective of the project to find cost savings and to find optimal strucuture alternatives for the European network.

But as experience from this project the scenarios could never have been calculated without using the above described modelling and data preparing tactics.

What could also be observed is that the main effect of the project results for the pratitioners had been far beyond the initally formulated objectives of reducing costs. The main effect had been the transparency of the network flows and the quantification of consequences as startpoint for new strategy discussions in the whole company.

4 Second Example: Application Project in the Food Industry

4.1 Description of the Project – Objectives

The second application example was performed in the food industry. The company in view is a producer of frozen food and is mainly active in Germany. In the status quo situation the distribution is done through three distribution centers which are located corresponding to the production sites of the company. The whole freight is operated by one freigth forwarder except for excess capacities which are carried out on the spot market.

The objective of the project had been the modelling of the very specific aspects of the status-quo situation as a baseline for comparison with scenarios. The disclosure of cost drivers and total cost had also been an objective. Moreover the scenario of one central warehouse for Germany should be estimated. Last not least the decision of choosing a new freight forwarder should be supported.

4.2 Coping with Shortcomings in Import Data Quality

In this application the quality of import data was not that bad. A simple preparation of the data formate combined with a standard data analysis was sufficient to create a realistic and plausible baseline for comparison of different scenario calculations.

The only problem in data quality was that the shipment data only contained pallet information instead of weight information.

To cope with this problem an analysis of the proportion of certain goods on pallets for special customers was carried out to find average weights per pallet due to customer numbers. The information of the shipment data base had then been

linked to the proportion results with article numbers as connecting element. On base of the information for each product and each customer a heuristic that converts the pallet information into weights could be built up in order to handle this convertion for mass data.

4.3 Coping with Modelling Limitations

Concerning the modelling limitations several very specific aspects had to be mapped (compare figure 5). For example were there direct interactions between certain customers. Caused by historically grown customer relations some key account customers had the convenience of being directly served by the company. These relations can't be queried.

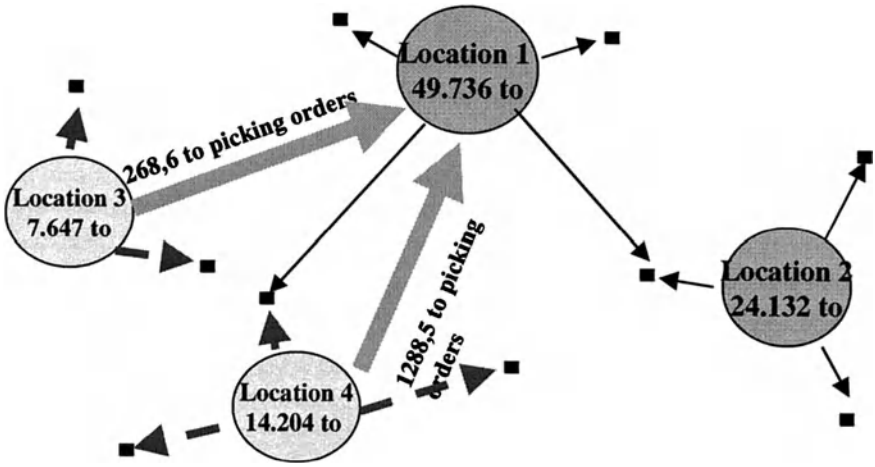


Fig. 5. Specific aspects of modeling historic grown customer relations

A second example for the very specific characteristics of the company's distribution policy is the existence of specific order picking rules. In some plant sites respectively distribution centers only a partial sortiment could be stored. A change into full sortiment locations would mean a lot of additional cost that also had to be modelled. In some plants only specific products could be produced as the food industry is bound to the local agricultural industry.

Last but not least the freigth forwarders being for chose had very special tariff systems with lots of exceptions. These exceptions mean for example that special areas in Germany are imposed with surcharges, or that pallets that overstep a certain weight must be divided in several shipments when calculating the prices.

To cope with the fixed direct interactions with customers the data is divided in several parts, each for the big key account customers and one for the „rest customer". Each data base is then modelled separately and put together only after the calculations.

The second tactic consists of splitting the model into several part models. As a result any order picking strategy could be mapped in a separate model and the consequences of each single strategy became obvious. The strategies in detail were a cross docking scenario, a scenario with one central warehouse, a full sortiment strategy in any distribution center and a mixed one.

To deal with the decision between the freight forwarders above all intensive discussions with the freigth forwarders took place. The question in focus was how the complicated tariffs with the lots of specific exceptions could be mapped in a weight and distance based tariff table. By eliminating the exceptions and translating them in regular table figures step by step a new tariff was generated. This one could easily be mapped in the tool but had nevertheless the acceptance of the freigth forwarders companies. Besides, this „tariff translation" helpd also the freigth forwarders reassuring their calculations and estimating the consequences of their complicated tariff exceptions.

4.4 Results

The results of this project had been unusual and surprising. The status quo situation of the three distribution centers with mixed strategy of full and part sortiment and no central distribution center had been the optimum. The practitioners had evolved the optimal situation for their company over the years and were consequently confirmed in their „feelings". Moreover they now had a written evidence of the rigthness of their strategies to hand over to their board of directors.

Another result had been the change of the freight forwarder. The transparency of cost especially in the jungle of tariff exceptions had shown up that the used forwarder had - compared to his service - much higher total costs. By the way had this also been the presumption of the practitioners at face. The experiences with the new freight forwarder up to now are very positive.

5 Conclusions and Considerations in the General Applicability of „Coping Strategies"

To sum up the experiences gained from the two application examples described above look at the following:

The biggest obstacle for efficient modelling in many cases is the preparation of a satisfactory data base. Connected with this demand is the creation of a baseline of the status quo that is understandable, tractable and plausible to the client. The task of preparing input data is the most time consuming, most difficult and least understood effort in many DSS-consulting projects.

A second conclusion is that not only cost savings are an important objective when modelling network optimizations. The transparency of the system and processes to the client, as well as the ability to illustrate its internal logic to the practioners is most critical. Only when the client can follow the logic in an interactive process of discussion and improvement between the modeller and the practicion-

ers there is a good chance for the development of mutually acceptable, truely plausible applications of standardized DSS tools. Frequently, the way to more transparency and traceability in modelling is splitting large data sets and very complex problems into sequences of partial problems that are solved incrementally .

A third conclusion from the experiences reported is, that more effort has to be expended in designing and testing coping heuristics for the revision and creation of appropriate data bases when the data provided directly from the client is not satisfactory. As a summary of the whole the following is a first attempt of formulating „coping strategies":

- Create missing data by projecting the existing information!
- Use proportions of given relevant parameters as baseline for heuristics!
- Analyze and quantify the status quo as much as possible and – discuss it with the practitioners!
- Whenever special relations or assignments are to be modeled – try to split them!

Notes:

1. For an application project and deeper details concerning the cooperation aspect as well as the tool itself see: Feige, d.; Klaus, P.; Werr, H.: „Decision Support for Designing Cooperative Distribution Networks" in:. M. Grazia Speranza, Paul Stähly (eds.): „New Trends in Distribution Logistics", IWDL 1999: Springer-Verlag Berlin Heidelberg, 1999, S. 63-93
2. An extremely useful, classical discussion of this idea is Lindblom, C.E.: „Still muddling, not yet through", in: Administrative Revue 1979, 39, S. 517-526

A Greedy Heuristic for a Three-level Multi-period Single-sourcing Problem

H. Edwin Romeijn*[1] and Dolores Romero Morales[2]

[1] Department of Industrial and Systems Engineering, University of Florida, 303 Weil Hall, P.O. Box 116595, Gainesville, Florida 32611-6595;
email: romeijn@ise.ufl.edu.

[2] Faculty of Economics and Business Administration, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands;
e-mail: d.romero@ke.unimaas.nl.

Abstract. In this paper we consider a model for integrating transportation and inventory decisions in a three-level logistics network consisting of plants, warehouses, and retailers. Our model includes production and throughput capacity constraints, and minimizes production, holding, and transportation costs in a dynamic environment. We show that the problem can be reformulated as a certain type of assignment problem with convex objective function. Based on this observation, we propose a greedy heuristic for the problem, and illustrate its behaviour on a class of randomly generated problem instances. These experiments suggest that the heuristic may be asymptotically feasible and optimal with probability one as the number of retailers increases.

Keywords: Dynamic models; integration of production, inventory, and transportation; dynamic demand pattern; greedy heuristic.

1 Introduction

The tendency to move towards global supply chains, the shortening of the product life cycle, and fast technological changes force companies to consider redesigning their logistics networks. The majority of the quantitative models proposed in the literature for the tactical problem of evaluating (usually with respect to costs) the layout of a distribution network assume a static environment. Hence the adequacy of those models is limited to situations where, in particular, the demand pattern is stationary over time. In addition, inventory decisions cannot be supported using stationary models.

In this paper we study a multi-period single-sourcing problem (MPSSP) that can be used for evaluating logistics network designs with respect to costs in a dynamic environment. The logistics network consists of a set of plants, a set of warehouses, and a set of retailers, see Figure 1. Each of the retailers experiences a given demand for a single product in each period and for a

* The work of this author was supported in part by the National Science Foundation under Grant No. DMI-0085682.

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

fixed planning horizon, and has to be supplied on time (i.e., no early deliveries or backlogs are allowed). Furthermore, each retailer needs to be delivered by (i.e., assigned to) a unique warehouse in each period. We assume that each plant has known, finite, and possibly time-varying capacity. Similarly, we consider that each warehouse has known, finite, and possibly time-varying throughput capacity. We assume that each warehouse has essentially unlimited physical capacity. The decisions that need to be made are (i) production sites and quantities, (ii) assignment of retailers to facilities, and (iii) location and size of inventories.

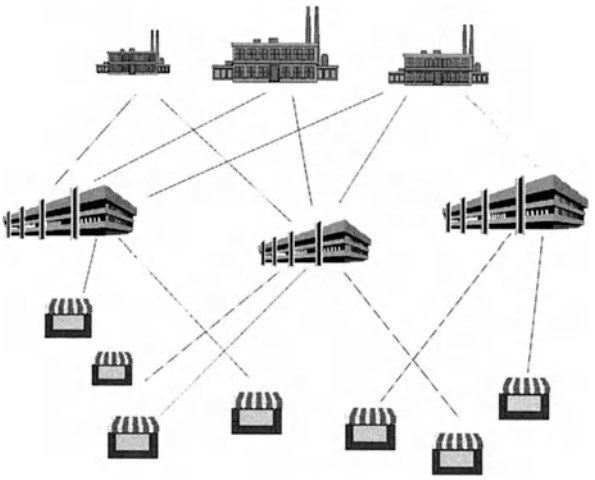


Fig. 1. Flows in the logistics distribution network

Romeijn and Romero Morales [14,15] have studied similar models where production and storage take place at the same location, and no throughput capacities are present. Due to the former assumption, the problem reduces to the evaluation of the design of a two-level logistics distribution network. The problem can then be formulated in terms of assignment variables only with an objective function that is separable in the warehouses. This reformulation then suggests a suitable class of greedy heuristics to find good feasible solutions for the problem. In contrast, when reformulating the three-level model that is the topic of this paper in terms of assignment variables only, the corresponding objective function is not separable in the warehouses. Nevertheless, we will be able to extend many of the results obtained for the two-level models to this larger class of problems. On the other hand, whereas for many two-level models it is possible to identify a greedy heuristic that is asymptotically optimal in a probabilistic sense as the number of retailers increases, we were not able to formally prove an analogous result for the greedy heuristic proposed in this paper for the three-level problem. However, partial results

as well as the numerical experiments lead us to conjecture that this result does extend to the three-level problem.

Since this problem is $\mathcal{NP}$-complete (see Martello and Toth [12] and Romero Morales, Van Nunen and Romeijn [17]), it is unlikely that efficient methods exist that can solve large problem instances to optimality. Therefore, it is appropriate to study heuristic approaches to this problem. We will show that our problem can be formulated as a certain type of assignment problem with convex objective function. This structure motivates the use of the class of greedy heuristics proposed by Martello and Toth [11] for the Generalized Assignment Problem (GAP), together with the family of pseudo-cost functions proposed by Romeijn and Romero Morales [14–16] for the GAP and two-level multi-period single-sourcing problems. Based on the structure of the LP-relaxation of our problem, we propose a suitable parameter choice, thereby identifying a particular greedy heuristic for the problem. We will provide numerical results on the performance of this heuristic, and conjecture that this member yields a heuristic that is asymptotically feasible and optimal in a probabilistic sense.

As mentioned above, related literature focuses mainly on static models. Examples of strategic/tactical models are Geoffrion and Graves [9], Benders et al. [2], and Fleischmann [7]. Duran [6] studies a dynamic model for the planning of production, bottling, and distribution of beer with an emphasis on the production process. Klose [10] analyzes the one-product version of the model proposed by Geoffrion and Graves [9]. Chan, Muriel and Simchi-Levi [3] study a dynamic, but uncapacitated, distribution problem in an operational setting. Arntzen et al. [1] present a multi-echelon multi-period model with no single-sourcing constraints on the assignment variables which was used in the reorganization of Digital Equipment Corporation.

The remainder of the paper is organized as follows. In Section 2 we will formulate the multi-period single-sourcing problem as a mixed-integer linear programming problem, and derive some properties of its LP-relaxation. In Section 3 we show the relationship with the GAP through a reformulation of the problem as a certain assignment problem with convex objective function. In Section 4 we will discuss a class of greedy heuristics for the problem, and select a suitable member of that class for which numerical experiments will be presented. The paper ends in Section 5 with some concluding remarks.

2 The Multi-period Single-sourcing Problem

2.1 A mixed-integer formulation

Let n denote the number of retailers, m the number of warehouses, q the number of plants, and T the length of the planning horizon. The demand of retailer j in period t is denoted by d_{jt}, while the production capacity at plant l in period t is equal to b_{lt}, and the maximal throughput capacity at warehouse i in period t is equal to r_{it}. The production, handling and transportation

costs per unit produced at plant l and transported to warehouse i in period t are c_{lit}. The costs of delivering the demand of retailer j from warehouse i in period t (i.e., the costs of assigning retailer j to warehouse i in period t) are a_{ijt}. Each retailer needs to be assigned to a single warehouse in any given period, which implies that the transportation costs can be an arbitrary (nonnegative) function of demand and distance. The inventory holding costs per unit at warehouse i in period t are h_{it}. (Note that all parameters are required to be nonnegative.)

The multi-period single-sourcing problem (MPSSP) can now be formulated as follows:

$$\text{minimize} \sum_{t=1}^{T}\sum_{l=1}^{q}\sum_{i=1}^{m} c_{lit}y_{lit} + \sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ijt}x_{ijt} + \sum_{t=1}^{T}\sum_{i=1}^{m} h_{it}I_{it}$$

subject to (P)

$$\sum_{j=1}^{n} d_{jt}x_{ijt} + I_{it} = \sum_{l=1}^{q} y_{lit} + I_{i,t-1}$$

$$i = 1,\ldots,m;\ t = 1,\ldots,T \tag{1}$$

$$\sum_{i=1}^{m} y_{lit} \le b_{lt} \tag{2}$$

$$l = 1,\ldots,q;\ t = 1,\ldots,T \tag{3}$$

$$\sum_{j=1}^{n} d_{jt}x_{ijt} \le r_{it} \tag{4}$$

$$i = 1,\ldots,m;\ t = 1,\ldots,T \tag{5}$$

$$I_{i0} = 0 \qquad i = 1,\ldots,m \tag{6}$$

$$\sum_{i=1}^{m} x_{ijt} = 1 \qquad j = 1,\ldots,n;\ t = 1,\ldots,T \tag{7}$$

$$x_{ijt} \in \{0,1\} \tag{8}$$

$$i = 1,\ldots,m;\ j = 1,\ldots,n;\ t = 1,\ldots,T \tag{9}$$

$$y_{lit} \ge 0 \qquad l = 1,\ldots,q;\ i = 1,\ldots,m;\ t = 1,\ldots,T$$

$$I_{it} \ge 0 \qquad i = 1,\ldots,m;\ t = 1,\ldots,T,$$

where y_{lit} is the amount produced at plant l and delivered to warehouse i in period t, x_{ijt} is equal to one if retailer j is assigned to warehouse i in period t and zero otherwise, and I_{it} denotes the inventory level at warehouse i at the end of period t. Constraints (1) impose the balance between the inflow, the storage and the outflow at warehouse i in period t. The maximal production capacity at plant l in period t is restricted by (3) and the maximal throughput capacity at warehouse i in period t by constraint (5). Without loss of generality, we impose in (6) that the inventory level at the beginning

of the planning horizon is equal to zero. Constraints (7) and (9) ensure that each retailer is delivered by exactly one warehouse in each period.

This model extends the well-known Single-Sourcing Problem (SSP) (see De Maio and Roveda [5]) in two directions. Firstly, the static character of the SSP prohibits the possibility of explicitly including decisions related to inventory management in the model. Secondly, the SSP assumes a layout of the distribution network where the production quantities are not included, or at least are not relevant (for instance when there is a one-to-one correspondence between warehouses and plants).

In the following section we will derive some properties of the LP-relaxation of the MPSSP. These properties will be useful when identifying a greedy heuristic for finding good feasible solutions for (P).

2.2 Properties of the LP-relaxation of the MPSSP

Denote the LP-relaxation of (P) by (LP) (see the Appendix). The following lemma derives an upper bound on the number of split assignments in any basic feasible solution for (LP). This upper bound is independent of the number of retailers. Let B be the set of (retailer-period)-pairs such that $(j,t) \in B$ means that retailer j is split in period t (i.e., retailer j is assigned to more than one warehouse in period t, each satisfying part of its demand).

Lemma 1. *If (LP) is feasible, any basic feasible solution of (LP) satisfies:*

$$|B| \leq 2mT + qT.$$

Proof. Rewrite the problem (LP) with equality constraints by introducing slack variables in the production capacity constraints (3) and in the throughput capacities constraints (5). We then obtain a problem with $mT + qT + mT + nT = 2mT + qT + nT$ equality constraints. Now consider a basic solution to (LP). The number of variables having a nonzero value in this solution is no larger than the number of equality constraints in the reformulated problem. Since there is *at least one* nonzero assignment variable corresponding to each assignment constraint (7), and *no more than one* nonzero assignment variable corresponding to each assignment that is feasible with respect to the integrality constraints of (P), the number of variables having a nonzero value is at least $|B| + nT$. Therefore, using the last two arguments we can derive that there can be no more than $2mT + qT$ assignments that are split.

After eliminating the variables I_{i0} using equations (16), the dual of (LP) can be formulated as

$$\text{maximize} \sum_{t=1}^{T} \sum_{j=1}^{n} v_{jt} - \sum_{t=1}^{T} \sum_{l=1}^{q} b_{lt}\omega_{lt} - \sum_{t=1}^{T} \sum_{i=1}^{m} r_{it}\nu_{it}$$

subject to (D)

$$
\begin{aligned}
v_{jt} &\leq a_{ijt} + \lambda_{it} d_{jt} + \nu_{it} d_{jt} && \\
& && i = 1, \ldots, m;\ j = 1, \ldots, n;\ t = 1, \ldots, T \\
\lambda_{it} &\leq \omega_{lt} + c_{lit} && \\
& && i = 1, \ldots, m;\ l = 1, \ldots, q;\ t = 1, \ldots, T \\
-\lambda_{it} + \lambda_{i,t+1} &\leq h_{it} && i = 1, \ldots, m;\ t = 1, \ldots, T-1 \\
\lambda_{it} &\quad \text{free} && i = 1, \ldots, m;\ t = 1, \ldots, T \\
\omega_{lt} &\geq 0 && l = 1, \ldots, q;\ t = 1, \ldots, T \\
\nu_{it} &\geq 0 && i = 1, \ldots, m;\ t = 1, \ldots, T \\
v_{jt} &\quad \text{free} && j = 1, \ldots, n;\ t = 1, \ldots, T.
\end{aligned}
$$

The following proposition suggests a way to use the dual optimal solution to distinguish split assignments from non-split ones. This result will be used in Section 4 to propose a class of greedy heuristics for (P).

Proposition 2. *Suppose that (LP) is feasible and non-degenerate. Let the vector* (x^*, y^*, I^*) *be a basic optimal solution for (LP) and let the vector* $(\lambda^*, \omega^*, \nu^*, v^*)$ *be the corresponding optimal solution for (D). Then,*

1. *For each* $(j,t) \notin B$, $x^*_{ijt} = 1$ *if and only if*

$$a_{ijt} + \lambda^*_{it} d_{jt} + \nu^*_{it} d_{jt} = \min_{k=1,\ldots,m} (a_{kjt} + \lambda^*_{kt} d_{jt} + \nu^*_{kt} d_{jt})$$

and

$$a_{ijt} + \lambda^*_{it} d_{jt} + \nu^*_{it} d_{jt} < \min_{k=1,\ldots,m;\ k \neq i} (a_{kjt} + \lambda^*_{kt} d_{jt} + \nu^*_{kt} d_{jt}).$$

2. *For each* $(j,t) \in B$, *there exists an index* i *such that*

$$a_{ijt} + \lambda^*_{it} d_{jt} + \nu^*_{it} d_{jt} = \min_{k=1,\ldots,m;\ k \neq i} (a_{kjt} + \lambda^*_{kt} d_{jt} + \nu^*_{kt} d_{jt}).$$

Proof. See the Appendix.

3 A Convex Assignment Formulation for the MPSSP

The MPSSP has been formulated as a mixed-integer linear programming problem in the assignment, production and inventory variables. The throughput constraints (5) together with constraints (7) and (9) suggest a relationship between the MPSSP and the GAP. In fact, we will prove in this section that (P) can be reformulated as a convex assignment problem in the variables x. The feasible region of this reformulation is formed by the Cartesian product of the feasible regions of T SSPs linked by the objective function, which is the sum of a linear and a convex function in x.

The following lemma will be used in the proof of Proposition 4. The lemma derives a necessary and sufficient condition for the feasibility of a dynamic extension of the standard transportation problem. As in the transportation problem, the problem is to satisfy the demand at a set of demand points from a set of supply points. However, the extension lies in the dynamic nature of the problem, and the fact that early deliveries to the warehouses are allowed. When applying the lemma, we will choose $\gamma_{lt} = b_{lt}$ and $\delta_{it} = \sum_{j=1}^{n} d_{jt} x_{ijt}$. The necessary and sufficient condition then says that the total demand in periods $1, \ldots, t$ should be no larger than the total capacity in periods $1, \ldots, t$, for all $t \in \{1, \ldots, T\}$.

Hereafter, let $\mathbb{R}_+$ denote the set of nonnegative real numbers, i.e., $\mathbb{R}_+ = [0, +\infty)$.

Lemma 3. *Let $\delta \in \mathbb{R}_+^{mT}$ and $\gamma \in \mathbb{R}_+^{qT}$. Then there exists a vector $y \in \mathbb{R}_+^{qmT}$ such that*

$$\sum_{i=1}^{m} y_{lit} \leq \gamma_{lt} \qquad \text{for all } l = 1, \ldots, q;\ t = 1, \ldots, T \tag{10}$$

and

$$\sum_{\tau=1}^{t} \sum_{l=1}^{q} y_{li\tau} \geq \sum_{\tau=1}^{t} \delta_{i\tau} \qquad \text{for all } i = 1, \ldots, m;\ t = 1, \ldots, T \tag{11}$$

if and only if

$$\sum_{\tau=1}^{t} \sum_{i=1}^{m} \delta_{i\tau} \leq \sum_{\tau=1}^{t} \sum_{l=1}^{q} \gamma_{l\tau} \qquad \text{for all } t = 1, \ldots, T. \tag{12}$$

Proof. See the Appendix.

Proposition 4. *Problem (P) can be reformulated as:*

$$\text{minimize} \sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ijt} x_{ijt} + H(x)$$

subject to *(P′)*

$$\sum_{j=1}^{n} d_{jt} x_{ijt} \leq r_{it} \qquad i = 1, \ldots, m;\ t = 1, \ldots, T$$

$$\sum_{i=1}^{m} x_{ijt} = 1 \qquad j = 1, \ldots, n;\ t = 1, \ldots, T$$

$$x_{ijt} \in \{0, 1\} \qquad i = 1, \ldots, m;\ j = 1, \ldots, n;\ t = 1, \ldots, T$$

where $H(x)$ is the convex function given by the optimal value of the following linear problem:

$$\text{minimize} \sum_{t=1}^{T}\sum_{l=1}^{q}\sum_{i=1}^{m} c_{lit}y_{lit} + \sum_{t=1}^{T}\sum_{i=1}^{m} h_{it}I_{it}$$

subject to

$$\begin{aligned}
I_{i,t-1} - I_{it} + \sum_{l=1}^{q} y_{lit} &= \sum_{j=1}^{n} d_{jt}x_{ijt} \\
&\qquad i = 1,\ldots,m;\ t = 1,\ldots,T \\
\sum_{i=1}^{m} y_{lit} &\le b_{lt} \qquad l = 1,\ldots,q;\ t = 1,\ldots,T \\
I_{i0} &= 0 \qquad i = 1,\ldots,m \\
y_{lit} &\ge 0 \qquad l = 1,\ldots,q;\ i = 1,\ldots,m;\ t = 1,\ldots,T \\
I_{it} &\ge 0 \qquad i = 1,\ldots,m;\ t = 1,\ldots,T.
\end{aligned}$$

Proof. The result follows by a decomposition argument. Let $\mathcal{F}$ be the feasible region of (P). We then have that

$$\begin{aligned}
&\min_{(x,y,I)\in\mathcal{F}} \left(\sum_{t=1}^{T}\sum_{l=1}^{q}\sum_{i=1}^{m} c_{lit}y_{lit} + \sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ijt}x_{ijt} + \right. \\
&\qquad \left. + \sum_{t=1}^{T}\sum_{i=1}^{m} h_{it}I_{it} \right) = \\
&= \min_{x:\exists(y',I')\,(x,y',I')\in\mathcal{F}} \left(\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ijt}x_{ijt} + \right. \\
&\qquad \left. \min_{(y,I):(x,y,I)\in\mathcal{F}} \left(\sum_{t=1}^{T}\sum_{l=1}^{q}\sum_{i=1}^{m} c_{lit}y_{lit} + \sum_{t=1}^{T}\sum_{i=1}^{m} h_{it}I_{it} \right) \right) \\
&= \min_{x:\exists(y',I')\,(x,y',I')\in\mathcal{F}} \left(\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ijt}x_{ijt} + H(x) \right).
\end{aligned}$$

Observe that

$$\sum_{\tau=1}^{t}\sum_{j=1}^{n} d_{j\tau} \le \sum_{\tau=1}^{t}\sum_{l=1}^{q} b_{l\tau} \qquad \text{for all } t = 1,\ldots,T \tag{13}$$

is a necessary condition for feasibility for both (P′) and the decomposed problem. Thus, hereafter we will assume that condition (13) holds. It remains to be shown that

$$\mathcal{F}' \equiv \{x \in \mathbb{R}^{mnT} \ : \ \exists\,(y,I) \in \mathbb{R}^{qmT} \times \mathbb{R}^{mT} \text{ such that } (x,y,I) \in \mathcal{F}\}$$

is the feasible region of (P′).

It is obvious that $\mathcal{F}'$ is contained in the feasible region of (P′). Now let x be a feasible vector for (P′). Lemma 3 and condition (13) imply that there exists a vector $y \in \mathbb{R}_+^{qmT}$ such that

$$\sum_{i=1}^{m} y_{lit} \leq b_{lt} \qquad l = 1, \ldots, q; t = 1, \ldots, T$$

and

$$\sum_{\tau=1}^{t} \sum_{l=1}^{q} y_{li\tau} \geq \sum_{\tau=1}^{t} \sum_{j=1}^{n} d_{j\tau} x_{ij\tau} \qquad i = 1, \ldots, m;\ t = 1, \ldots, T.$$

Now define I_{it}, for $i = 1, \ldots, m$ and $t = 1, \ldots, T$, as

$$I_{it} = \sum_{\tau=1}^{t} \sum_{l=1}^{q} y_{li\tau} - \sum_{\tau=1}^{t} \sum_{j=1}^{n} d_{j\tau} x_{ij\tau}.$$

It is easy to see that I_{it} is nonnegative and $(x, y, I) \in \mathcal{F}$, and thus $x \in \mathcal{F}'$.

Using strong duality for linear programming, it is straightforward to show that the function H is convex.

This result shows that, for each assignment solution to (P′), corresponding optimal values for the production and inventory variables exist. A similar result was derived by Freling et al. [8], for the case where there exists a one-to-one correspondence between warehouses and plants. In this case, the objective function of the assignment problem is separable in the index i. The separability of the objective function allows the reformulation of the problem as a set partitioning problem, which can be used to construct a Branch and Price algorithm for this class of problems.

4 Solving the MPSSP

4.1 A greedy heuristic for the MPSSP

In the previous section we have shown that the MPSSP can be formulated as a collection of T SSPs that are joined through a convex objective function. Since the SSP is a special case of the GAP, we propose to use a greedy heuristic similar to the one given by Martello and Toth [11] for the GAP, using a pseudo-cost function from the family introduced by Romeijn and Romero Morales [16].

The idea behind the greedy heuristic is that each possible assignment of a (retailer,period)-pair (j, t) to a warehouse i is evaluated by a pseudo-cost function $f(i, j, t)$. For each assignment to be made, the difference between the two smallest values of $f(i, j, t)$ (called the *desirability* of making the cheapest

assignment with respect to the pseudo-cost) is computed, and assignments are made in decreasing order of this difference. Along the way, the remaining capacities of the warehouses, and consequently the values of the desirabilities, are updated to ensure feasibility. Note, from formulation (P') of the MPSSP, that only the throughput capacities play a role with respect to feasibility.

Romeijn and Romero Morales [14,15] propose to use the following family of pseudo-cost functions:

$$f_\alpha(i,j,t) = a_{ijt} + \alpha_{it} d_{jt}$$

where $\alpha \in \mathbb{R}_+^{mT}$. We may observe that this pseudo-cost function combines costs (a_{ijt}) with demands (d_{jt}) (i.e., the use of the scarce throughput capacity at the warehouses). For the two-level MPSSP, an analogous result to Proposition 2 suggests choosing α to be essentially the vector of dual multipliers to the production capacity constraints. In fact, for many classes of the MPSSP this choice is asymptotically optimal in a probabilistic sense if $n \to \infty$.

For the three-level model considered in this paper, the result of Proposition 2, where the split and nonsplit assignments in the LP-relaxation of (P) are characterized, suggests in a similar manner to use the member of the family of pseudo-cost functions given by

$$\alpha_{it} = \lambda_{it}^* + \nu_{it}^*$$

(yielding $f(i,j,t) = a_{ijt} + (\lambda_{it}^* + \nu_{it}^*) d_{jt}$), where $\lambda^* \in \mathbb{R}_+^{mT}$ is the vector of optimal dual multipliers of the flow conservation constraints in (LP), and similarly $\nu^* \in \mathbb{R}_+^{mT}$ is the vector of optimal dual multipliers of the throughput constraints (where the corresponding constraints are reformulated as $\geq$-constraints, so that the dual multipliers are nonnegative). As in the two-level case, it may seem surprising at first sight that the inventory holding costs are not explicitly incorporated into the pseudo-cost function. However, note that the choice of α as a function of the optimal dual multipliers of the capacity constraints accounts implicitly for these costs through the dual constraints that are clearly satisfied by these multipliers.

Proposition 2 ensures that the heuristic starts by making assignments which are feasible for (LP) with respect to the integrality constraints. In a similar way as in Romeijn and Romero Morales [14,15], we can show that the number of differences between the feasible assignments of (LP) and the assignments of the greedy heuristic is bounded by a constant independent of the number of retailers. Therefore, the assignments given by the greedy heuristic are close to the assignments of (LP). Since the number of infeasible assignments in (LP) is again bounded by a constant independent of the number of retailers by Lemma 1, we expect that the greedy solution is close to an optimal integer solution. A probabilistic analysis of such heuristics on similar problems leads us to conjecture that this choice will yield a heuristic that is asymptotically optimal in a probabilistic sense (as n goes to ∞). The fact that the objective function of the reformulation in the assignment variables

is not separable in the warehouses has not allowed us to formally prove this conjecture.

4.2 Some numerical results

In this section we will illustrate the behaviour of the greedy heuristic as described in the previous section on a set of randomly generated test problems. For each problem instance, we generate a set of n retailers, a set of m warehouses, and a set of q plants uniformly in the square $[0, 10]^2$. For retailer j ($j = 1, \ldots, n$), we generate a random demand D_{jt} in period t ($t = 1, \ldots, T$) from the uniform distribution on $[5\sigma_t, 25\sigma_t]$, where the vector σ contains seasonal factors, which we have chosen to be $\sigma = (\frac{1}{2}, \frac{3}{4}, 1, 1, \frac{3}{4}, \frac{1}{2})^\top$. The production costs are assumed to be equal to the distance, i.e., $c_{lit} = \text{dist}_{li}$, where dist_{li} denotes the Euclidean distance between plant l and warehouse i. The assignment costs are assumed to be proportional to demand and distance, i.e., $a_{ijt} = d_{jt} \cdot \text{dist}_{ij}$, where dist_{ij} denotes the Euclidean distance between warehouse i and retailer j. Finally, we generate inventory holding costs H_{it} uniformly from $[10, 30]$.

We have chosen the capacities equal to $b_{lt} = \frac{1}{q} \cdot \beta \cdot n$ and $r_{it} = \frac{1}{m} \cdot \rho \cdot n$, where

$$\beta = \delta \cdot 15 \cdot \max_{t=1,\ldots,T} \left(\frac{1}{t} \sum_{\tau=1}^{t} \sigma_\tau \right)$$

$$\rho = \delta \cdot 15 \cdot \max_{t=1,\ldots,T} \sigma_t.$$

The results of Lemma 3 and Romeijn and Piersma [13] show that the instances generated by this probabilistic model are asymptotically feasible with probability one (as n goes to ∞) if $\delta > 1$, and infeasible with probability one (again as n goes to ∞) if $\delta < 1$. To account for the asymptotic nature of this feasibility guarantee, we have set $\delta = 1.1$ to obtain feasible instances for finite n.

We have chosen the number of plants equal to $q = 3, 4, 5$, the number of warehouses equal to $m = 5, 10, 15$, and the number of periods equal to $T = 6, 12$ and 18. We let the number of retailers vary from $n = 50$ until $n = 500$ in increments of 50 retailers. For each size of the problem we have generated 50 instances. All the runs were performed on a PC with a 866 MHz Pentium III processor and 128 MB RAM. All instances of (LP) were solved using CPLEX 6.6 [4].

Tables 1–9 illustrate the behaviour of the greedy heuristic (using the pseudo-cost function mentioned in Section 4.1). Clearly, n denotes the number of retailers. Each table shows the number of instances for which the LP-relaxation was feasible, as well as the number of instances for which the heuristic found a feasible solution. In addition, the time needed to solve the LP-relaxation, as well as the time employed by the heuristic, not including the

time needed to solve the LP-relaxation, is shown. Finally, an upper bound on the average error of the heuristic solution is shown, as measured by the relative deviation of the heuristic solution value from the optimal LP-value. This average was calculated only using the instances where the heuristic found a feasible solution.

	LP						heuristic								
	# feasible			time (sec.)			# feasible			time (sec.)			error (%)		
n T	6	12	18	6	12	18	6	12	18	6	12	18	6	12	18
50	42	50	44	0.08	0.22	0.37	37	43	40	0.01	0.01	0.02	0.99	0.69	0.63
100	48	50	48	0.19	0.69	1.23	48	49	48	0.01	0.02	0.07	0.47	0.36	0.30
150	50	50	50	0.33	1.12	2.00	50	50	50	0.01	0.06	0.07	0.33	0.23	0.21
200	50	50	50	0.59	1.63	3.21	50	50	50	0.04	0.06	0.07	0.20	0.16	0.15
250	50	50	50	0.81	2.43	4.96	50	50	50	0.05	0.05	0.06	0.17	0.13	0.14
300	50	50	50	0.97	3.20	6.19	50	50	50	0.05	0.05	0.06	0.17	0.14	0.11
350	50	50	50	1.31	4.41	9.68	50	50	50	0.05	0.06	0.06	0.11	0.10	0.09
400	50	50	50	1.51	5.59	11.66	50	50	50	0.06	0.05	0.07	0.10	0.09	0.07
450	50	50	50	1.71	7.06	15.67	50	50	50	0.06	0.05	0.08	0.09	0.08	0.07
500	50	50	50	2.31	9.02	18.50	50	50	50	0.05	0.05	0.09	0.08	0.07	0.06

Table 1. Greedy heuristic when $q = 3$ and $m = 5$

Although we cannot guarantee that the heuristic will always find a feasible solution (recall that even to determine whether a particular instance of the MPSSP is feasible is an $\mathcal{NP}$-complete problem), a feasible solution was always found for instances with at least 300 retailers for the different values of q, m and T considered. Observe that for all instances with 150, 200 and 250 retailers, a total 4050 instances, the heuristic just failed to find a feasible solution in 8. Note that feasibility of the LP-relaxation does not imply feasibility of the MPSSP, so that the inability of the heuristic to find a feasible solution could be caused by infeasibility of the instance, even when the LP-relaxation is feasible.

The average error was always well below 1.1% for $n \geq 150$, and the quality of the greedy solution improves with increasing value of T. Generally speaking, the upper bound on the error increases with the number of warehouses. The number of plants does not seem to have a significant effect on the quality of the solution obtained by the heuristic. Moreover, the fact that the average error decreases as the number of retailers increases supports our conjecture that the heuristic is asymptotically optimal.

With respect to the computation times, we observe that the most time is consumed in finding the optimal dual multipliers of (LP) which are needed to define our choice of the pseudo-cost function. The time required to solve (LP) clearly increases with n and T, but is very reasonable.

	LP						heuristic								
	# feasible			time (sec.)			# feasible			time (sec.)			error (%)		
n T	6	12	18	6	12	18	6	12	18	6	12	18	6	12	18
50	45	45	44	0.20	0.70	1.18	33	34	35	0.02	0.02	0.08	2.28	1.68	1.40
100	50	49	50	0.58	1.51	2.57	48	49	49	0.03	0.07	0.11	1.01	0.74	0.71
150	50	50	49	0.87	2.39	4.59	50	50	49	0.06	0.07	0.10	0.70	0.50	0.45
200	50	50	50	1.30	3.78	6.65	50	50	50	0.06	0.07	0.12	0.53	0.39	0.33
250	50	50	50	1.80	5.00	10.28	50	50	50	0.06	0.07	0.12	0.39	0.31	0.26
300	50	50	50	2.16	6.67	13.81	50	50	50	0.06	0.08	0.13	0.30	0.24	0.21
350	50	50	50	2.66	9.04	20.41	50	50	50	0.06	0.07	0.14	0.26	0.20	0.18
400	50	50	50	3.33	10.57	26.12	50	50	50	0.06	0.08	0.19	0.23	0.21	0.17
450	50	50	50	4.09	14.62	30.94	50	50	50	0.05	0.09	0.20	0.21	0.15	0.15
500	50	50	50	5.15	18.48	38.04	50	50	50	0.06	0.13	0.21	0.20	0.16	0.15

Table 2. Greedy heuristic when $q = 3$ and $m = 10$

	LP						heuristic								
	# feasible			time (sec.)			# feasible			time (sec.)			error (%)		
n T	6	12	18	6	12	18	6	12	18	6	12	18	6	12	18
50	47	44	44	0.33	1.02	1.82	26	24	18	0.02	0.09	0.15	3.93	2.96	2.24
100	50	48	50	0.85	2.43	4.05	46	46	48	0.06	0.10	0.19	1.69	1.27	1.10
150	50	50	50	1.34	3.93	6.95	50	50	48	0.06	0.11	0.21	1.08	0.81	0.66
200	50	50	50	2.02	5.01	9.63	50	50	50	0.07	0.10	0.21	0.76	0.53	0.52
250	50	50	50	2.44	7.76	15.14	50	50	50	0.07	0.10	0.22	0.63	0.48	0.42
300	50	50	50	3.34	10.43	20.52	50	50	50	0.07	0.10	0.29	0.51	0.40	0.32
350	50	50	50	4.07	14.75	30.92	50	50	50	0.06	0.15	0.30	0.46	0.31	0.26
400	50	50	50	5.56	16.76	31.88	50	50	50	0.06	0.18	0.32	0.33	0.29	0.24
450	50	50	50	6.36	21.22	42.61	50	50	50	0.06	0.18	0.37	0.29	0.26	0.21
500	50	50	50	7.77	24.77	52.42	50	50	50	0.07	0.19	0.42	0.32	0.22	0.21

Table 3. Greedy heuristic when $q = 3$ and $m = 15$

	LP						heuristic								
	# feasible			time (sec.)			# feasible			time (sec.)			error (%)		
n T	6	12	18	6	12	18	6	12	18	6	12	18	6	12	18
50	47	48	44	0.09	0.22	0.39	40	43	38	0.01	0.01	0.02	1.13	0.61	0.53
100	50	49	50	0.19	0.69	1.29	49	47	49	0.02	0.04	0.06	0.47	0.32	0.29
150	49	49	50	0.35	1.14	2.17	49	49	50	0.02	0.05	0.06	0.31	0.21	0.21
200	50	50	50	0.57	1.73	3.13	50	50	50	0.06	0.06	0.07	0.26	0.16	0.17
250	50	50	50	0.80	2.66	5.28	50	50	50	0.06	0.06	0.07	0.19	0.13	0.12
300	50	50	50	1.05	3.46	7.07	50	50	50	0.06	0.06	0.06	0.16	0.11	0.10
350	50	50	50	1.35	4.57	10.34	50	50	50	0.05	0.06	0.06	0.15	0.09	0.08
400	50	50	50	1.53	6.11	13.54	50	50	50	0.05	0.06	0.08	0.12	0.07	0.06
450	50	50	50	1.90	8.12	18.17	50	50	50	0.05	0.05	0.07	0.10	0.08	0.07
500	50	50	50	2.28	9.76	21.06	50	50	50	0.06	0.05	0.10	0.08	0.06	0.05

Table 4. Greedy heuristic when $q = 4$ and $m = 5$

	LP						heuristic								
	# feasible			time (sec.)			# feasible			time (sec.)			error (%)		
n T	6	12	18	6	12	18	6	12	18	6	12	18	6	12	18
50	41	43	49	0.20	0.77	1.34	32	34	31	0.01	0.03	0.09	2.53	1.71	1.56
100	49	49	50	0.58	1.69	3.02	45	46	47	0.04	0.07	0.10	1.10	0.95	0.75
150	50	50	50	1.02	2.86	5.21	50	50	50	0.06	0.07	0.11	0.77	0.54	0.50
200	50	50	50	1.35	4.32	8.00	50	49	50	0.05	0.07	0.11	0.53	0.38	0.33
250	50	50	50	1.98	5.66	10.86	50	50	50	0.07	0.08	0.13	0.40	0.30	0.29
300	50	50	50	2.39	8.16	16.43	50	50	50	0.05	0.07	0.13	0.36	0.27	0.23
350	50	50	50	3.29	11.91	21.45	50	50	50	0.06	0.07	0.13	0.25	0.23	0.18
400	50	50	50	3.59	14.21	27.71	50	50	50	0.07	0.08	0.21	0.24	0.21	0.18
450	50	50	50	4.57	17.47	36.87	50	50	50	0.06	0.09	0.21	0.22	0.18	0.15
500	50	50	50	5.38	24.31	49.32	50	50	50	0.04	0.13	0.22	0.20	0.15	0.13

Table 5. Greedy heuristic when $q = 4$ and $m = 10$

	LP						heuristic								
	# feasible			time (sec.)			# feasible			time (sec.)			error (%)		
n T	6	12	18	6	12	18	6	12	18	6	12	18	6	12	18
50	45	47	46	0.33	1.16	1.99	25	23	23	0.02	0.09	0.17	3.82	3.15	2.55
100	50	50	50	0.99	2.52	4.84	47	47	46	0.06	0.10	0.18	1.61	1.24	1.22
150	50	50	50	1.51	4.24	8.51	49	50	49	0.06	0.10	0.21	1.09	0.92	0.79
200	50	50	50	2.25	6.63	11.75	50	49	50	0.06	0.09	0.22	0.83	0.62	0.52
250	50	50	50	2.94	9.35	16.42	50	50	50	0.06	0.10	0.22	0.64	0.46	0.44
300	50	50	50	3.87	11.60	25.77	50	50	50	0.07	0.11	0.29	0.52	0.38	0.34
350	50	50	50	5.00	15.71	34.64	50	50	50	0.07	0.13	0.31	0.44	0.34	0.30
400	50	50	50	6.15	21.56	45.72	50	50	50	0.05	0.18	0.33	0.41	0.30	0.27
450	50	50	50	7.37	25.16	54.47	50	50	50	0.06	0.18	0.33	0.37	0.25	0.21
500	50	50	50	9.35	30.07	64.95	50	50	50	0.07	0.19	0.35	0.32	0.26	0.21

Table 6. Greedy heuristic when $q = 4$ and $m = 15$

	LP						heuristic								
	# feasible			time (sec.)			# feasible			time (sec.)			error (%)		
n T	6	12	18	6	12	18	6	12	18	6	12	18	6	12	18
50	46	47	49	0.10	0.25	0.66	38	43	44	0.00	0.01	0.02	0.99	0.78	0.63
100	50	49	49	0.23	0.85	1.46	50	49	49	0.01	0.03	0.04	0.52	0.34	0.31
150	50	50	50	0.39	1.49	2.69	50	50	50	0.03	0.06	0.05	0.28	0.24	0.19
200	50	50	50	0.69	2.10	4.52	50	50	50	0.06	0.06	0.07	0.21	0.18	0.17
250	49	50	50	0.97	3.20	6.87	49	50	50	0.06	0.05	0.08	0.18	0.13	0.11
300	50	50	50	1.28	4.43	8.60	50	50	50	0.06	0.06	0.08	0.13	0.12	0.11
350	50	50	50	1.74	5.75	13.41	50	50	50	0.04	0.06	0.10	0.14	0.10	0.09
400	50	50	50	2.01	7.94	16.89	50	50	50	0.06	0.04	0.11	0.13	0.09	0.08
450	50	50	50	2.38	10.25	22.77	50	50	50	0.05	0.06	0.11	0.10	0.08	0.07
500	50	50	50	2.93	13.34	27.84	50	50	50	0.05	0.08	0.12	0.08	0.08	0.05

Table 7. Greedy heuristic when $q = 5$ and $m = 5$

	LP						heuristic								
	# feasible			time (sec.)			# feasible			time (sec.)			error (%)		
n T	6	12	18	6	12	18	6	12	18	6	12	18	6	12	18
50	45	41	45	0.21	0.77	1.48	29	26	29	0.03	0.03	0.08	2.48	1.95	1.71
100	49	50	50	0.74	1.92	3.16	47	50	48	0.02	0.05	0.11	1.21	0.79	0.80
150	50	50	50	1.15	3.14	6.12	50	50	50	0.03	0.07	0.12	0.69	0.57	0.51
200	50	50	50	1.67	4.59	10.30	50	50	50	0.02	0.07	0.13	0.52	0.41	0.37
250	49	50	50	2.31	6.66	13.71	49	50	50	0.04	0.09	0.16	0.39	0.33	0.30
300	50	50	50	2.92	9.22	18.48	50	50	50	0.04	0.09	0.17	0.33	0.26	0.23
350	50	50	50	3.64	12.55	24.37	50	50	50	0.05	0.11	0.18	0.29	0.23	0.20
400	50	50	50	4.47	14.89	33.67	50	50	50	0.06	0.11	0.21	0.25	0.18	0.14
450	50	50	50	5.15	19.81	44.26	50	50	50	0.08	0.13	0.24	0.23	0.17	0.16
500	50	50	50	6.73	24.54	55.87	50	50	50	0.09	0.14	0.24	0.17	0.17	0.15

Table 8. Greedy heuristic when $q = 5$ and $m = 10$

	LP						heuristic								
	# feasible			time (sec.)			# feasible			time (sec.)			error (%)		
n T	6	12	18	6	12	18	6	12	18	6	12	18	6	12	18
50	46	47	48	0.47	1.32	2.27	20	27	24	0.03	0.07	0.16	4.14	3.06	2.73
100	50	50	50	1.10	2.83	5.20	47	46	50	0.05	0.10	0.18	1.61	1.43	1.27
150	50	50	50	1.86	4.59	8.20	49	49	50	0.04	0.10	0.24	1.05	0.87	0.77
200	50	50	50	2.62	6.97	15.93	50	50	50	0.05	0.12	0.25	0.83	0.63	0.54
250	50	50	50	3.26	9.69	19.27	49	50	50	0.06	0.14	0.26	0.68	0.48	0.45
300	50	50	50	4.22	12.27	26.04	50	50	50	0.07	0.16	0.30	0.51	0.45	0.37
350	50	50	50	5.35	17.33	36.52	50	50	50	0.08	0.17	0.35	0.40	0.32	0.31
400	50	50	50	6.92	23.10	52.52	50	50	50	0.10	0.18	0.46	0.37	0.28	0.27
450	50	50	50	8.12	27.26	57.16	50	50	50	0.10	0.20	0.60	0.34	0.27	0.24
500	50	50	50	9.77	35.94	71.41	50	50	50	0.10	0.23	0.62	0.32	0.24	0.22

Table 9. Greedy heuristic when $q = 5$ and $m = 15$

In addition to using the heuristic, we have also called the MIP solver of CPLEX to try to solve small instances of the problem to optimality. We have considered $q = 3$, $m = 5$, $n = 50, 100$, and $T = 6, 12, 18$. The procedure was stopped after 30 minutes of CPU time. In Table 10, we present the number of times that the MIP solver of CPLEX was able to prove optimality within 30 minutes. We also give the average time employed in these instances, and the upper bound on the error on the instances in which the MIP solver of CPLEX could not prove optimality. Although the MIP solver of CPLEX did always find a feasible solution to all feasible instances within 30 minutes, the time needed to find an optimal solution is extremely large. We conclude that the heuristic is a very effective way of finding a high quality solution with little effort.

	# solved			time (sec.)			error(%)		
n T	6	12	18	6	12	18	6	12	18
50	41	27	19	141.30	171.05	157.21	0.44	0.33	0.28
100	31	16	9	417.95	750.25	167.17	0.12	0.13	0.12

Table 10. CPLEX for small instances when $q = 3$ and $m = 5$

5 Conclusions

In this paper we have analyzed a model for evaluating the design of a logistics network in a dynamic environment. The network consists of plants, warehouses and retailers. The model deals with production and throughput constraints, as well as standard single-sourcing constraints. Based on a reformulation of the problem as a convex assignment problem, we have proposed a greedy heuristic. The numerical illustrations indicate that the heuristic may be asymptotically feasible and optimal.

References

1. B.C. Arntzen, G.G. Brown, T.P. Harrison, and L.L. Trafton. Global supply chain management at Digital Equipment Corporation. *Interfaces*, 25(1):69–93, 1995.
2. J.F. Benders, W.K.M. Keulemans, J.A.E.E. van Nunen, and G. Stolk. A decision support program for planning locations and allocations with the aid of linear programming. In C.B. Tilanus, O.B. de Gaus, and J.K. Lenstra, editors, *Quantitative Methods in Management: cases studies of failures and successes*, chapter 4, pages 29–34. John Wiley & Sons, Chichester, 1986.

3. L.M.A. Chan, A. Muriel, and D. Simchi-Levi. Production/distribution planning problems with piece-wise linear and concave transportation costs. Research Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois, 1998.

4. CPLEX Reference Manual. *ILOG CPLEX 6.6.* ILOG, Inc., Incline Village, Nevada, 1999.

5. A. De Maio and C. Roveda. An all zero-one algorithm for a certain class of transportation problems. *Operations Research*, 19(6):1406–1418, 1971.

6. F. Duran. A large mixed integer production and distribution program. *European Journal of Operational Research*, 28:207–217, 1987.

7. B. Fleischmann. Designing distribution systems with transport economies of scale. *European Journal of Operational Research*, 70:31–42, 1993.

8. R. Freling, H.E. Romeijn, D. Romero Morales, and A.P.M. Wagelmans. A branch and price algorithm for the multi-period single-sourcing problem. ERASM Management Report Series no. 49-1999, Rotterdam School of Management, Erasmus University Rotterdam, 1999.

9. A.M. Geoffrion and G.W. Graves. Multicommodity distribution system design by Benders decomposition. *Management Science*, 20(5):822–844, 1974.

10. A. Klose. An LP-based heuristic for two-stage capacitated facility location problems. *Journal of the Operational Research Society*, 50:157–166, 1999.

11. S. Martello and P. Toth. An algorithm for the generalized assignment problem. In J.P. Brans, editor, *Operational Research '81*, pages 589–603. IFORS, North–Holland, Amsterdam, 1981.

12. S. Martello and P. Toth. *Knapsack problems, algorithms and computer implementations.* John Wiley & Sons, New York, 1990.

13. H.E. Romeijn and N. Piersma. A probabilistic feasibility and value analysis of the generalized assignment problem. *Journal of Combinatorial Optimization*, 4(3):325–355, 2000.

14. H.E. Romeijn and D. Romero Morales. Asymptotic analysis of a greedy heuristic for the multi-period single-sourcing problem: the acyclic case. Research Report 99-13, Department of Industrial and Systems Engineering, University of Florida, 1999.

15. H.E. Romeijn and D. Romero Morales. An asymptotically optimal greedy heuristic for the multi-period single-sourcing problem: the cyclic case. ERASM Management Report Series no. 20-1999, Rotterdam School of Management, Erasmus University Rotterdam, 1999.

16. H.E. Romeijn and D. Romero Morales. A class of greedy algorithms for the generalized assignment problem. *Discrete Applied Mathematics*, 103:209–235, 2000.

17. D. Romero Morales, J.A.E.E. van Nunen, and H.E. Romeijn. Logistics network design evaluation in a dynamic environment. In M.G. Speranza and P. Stähly, editors, *New trends in distribution logistics, Lecture notes in economics and mathematical systems 480*, pages 113–135. Springer-Verlag, Berlin, 1999.

Appendix

The LP-relaxation of (P) can be formulated as follows:

$$\text{minimize} \sum_{t=1}^{T}\sum_{l=1}^{q}\sum_{i=1}^{m} c_{lit}y_{lit} + \sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ijt}x_{ijt} + \sum_{t=1}^{T}\sum_{i=1}^{m} h_{it}I_{it}$$

subject to (LP)

$$\sum_{j=1}^{n} d_{jt}x_{ijt} + I_{it} = \sum_{l=1}^{q} y_{lit} + I_{i,t-1} \qquad i = 1,\ldots,m;\ t = 1,\ldots,T$$

$$\sum_{i=1}^{m} y_{lit} \le b_{lt} \qquad l = 1,\ldots,q;\ t = 1,\ldots,T \tag{14}$$

$$\sum_{j=1}^{n} d_{jt}x_{ijt} \le r_{it} \qquad i = 1,\ldots,m;\ t = 1,\ldots,T \tag{15}$$

$$I_{i0} = 0 \qquad i = 1,\ldots,m \tag{16}$$

$$\sum_{i=1}^{m} x_{ijt} = 1 \qquad j = 1,\ldots,n;\ t = 1,\ldots,T \tag{17}$$

$$x_{ijt} \ge 0 \qquad i = 1,\ldots,m;\ j = 1,\ldots,n;\ t = 1,\ldots,T$$

$$y_{lit} \ge 0 \qquad l = 1,\ldots,q;\ i = 1,\ldots,m;\ t = 1,\ldots,T$$

$$I_{it} \ge 0 \qquad i = 1,\ldots,m;\ t = 1,\ldots,T.$$

Let (x^*, y^*, I^*) be a basic optimal solution for (LP). In the following lemma, which will be used in the proof of Proposition 2, we derive a relationship between the number of split assignments, the number of fractional assignment variables, the number of (plant,warehouse,period)-triples having a positive flow, the number of (plant,period)-pairs where the plant is used to full capacity in that period, the number of (warehouse,period)-pairs where the warehouse is used to full capacity in that period and the number of strictly positive inventory variables. Let F be the set of fractional assignment variables, Q the set of (plant,period)-pairs where the plant is used to full capacity in that period, W the set of (warehouse,period)-pairs where the warehouse is used to full capacity in that period, Y^+ the set of (plant,warehouse,period)-triples having a positive flow and I^+ the set of strictly positive inventory variables, i.e.

$$F = \{(i,j,t) \,:\, 0 < x^*_{ijt} < 1\}$$

$$Q = \{(l,t) \,:\, \sum_{i=1}^{m} y^*_{lt} = b_{lt}\}$$

$$W = \{(i,t) : \sum_{j=1}^{n} d_{jt} x^*_{ijt} = r_{it}\}$$
$$Y^+ = \{(l,i,t) : y^*_{lit} > 0\}$$
$$I^+ = \{(i,t) : I^*_{it} > 0\}.$$

Lemma A.1. *If (LP) is feasible and non-degenerate, then for a basic optimal solution of (LP) we have*

$$|F| + |Y^+| + |I^+| = mT + |Q| + |W| + |B|.$$

Proof. Denote by s_{lt} the slack variables corresponding to the production capacity constraints in (LP) and S_{it} the slack variables corresponding to the throughput capacity constraints. Thus, including these variables, (LP) can be reformulated as

$$\text{minimize} \sum_{t=1}^{T}\sum_{l=1}^{q}\sum_{i=1}^{m} c_{lit} y_{lit} + \sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ijt} x_{ijt} + \sum_{t=1}^{T}\sum_{i=1}^{m} h_{it} I_{it}$$

subject to

$$\sum_{j=1}^{n} d_{jt} x_{ijt} + I_{it} = \sum_{l=1}^{q} y_{lit} + I_{i,t-1} \qquad i = 1,\ldots,m;\ t = 1,\ldots,T$$
$$\sum_{i=1}^{m} y_{lit} + s_{lt} = b_{lt} \qquad l = 1,\ldots,q;\ t = 1,\ldots,T$$
$$\sum_{j=1}^{n} d_{jt} x_{ijt} + S_{it} = r_{it} \qquad i = 1,\ldots,m;\ t = 1,\ldots,T$$
$$I_{i0} = 0 \qquad i = 1,\ldots,m$$
$$\sum_{i=1}^{m} x_{ijt} = 1 \qquad j = 1,\ldots,n;\ t = 1,\ldots,T$$
$$x_{ijt} \geq 0 \qquad i = 1,\ldots,m;\ j = 1,\ldots,n;\ t = 1,\ldots,T$$
$$y_{lit} \geq 0 \qquad l = 1,\ldots,q;\ i = 1,\ldots,m;\ t = 1,\ldots,T$$
$$I_{it} \geq 0 \qquad i = 1,\ldots,m;\ t = 1,\ldots,T$$
$$s_{lt} \geq 0 \qquad l = 1,\ldots,q;\ t = 1,\ldots,T$$
$$S_{it} \geq 0 \qquad i = 1,\ldots,m;\ t = 1,\ldots,T.$$

Let $(x^*, y^*, I^*, s^*, S^*)$ be a basic optimal solution for (LP). Then, sets Q and W, defined above, are equal to

$$Q = \{(l,t) : s^*_{lt} = 0\}$$
$$W = \{(i,t) : S^*_{it} = 0\}.$$

Under non-degeneracy, the number of nonzero variables at $(x^*, y^*, I^*, s^*, S^*)$ is equal to $2mT + qT + nT$, the number of constraints in (LP). The number of nonzero assignment variables is equal to $(nT - |B|) + |F|$, where the first term corresponds to the variables $x^*_{ijt} = 1$, the second one to the fractional assignment variables. With respect to the slack variables, we have $(qT - |Q|) + (mT - |W|)$ nonzero variables. By definition $|Y^+|$ is the number of nonzero production variables. The same follows for I^+. Thus, by imposing that the number of nonzero variables at $(x^*, y^*, I^*, s^*, S^*)$ is equal to $2mT + qT + nT$, we obtain

$$\begin{aligned} &2mT + qT + nT = \\ &= (nT - |B|) + |F| + (qT - |Q|) + (mT - |W|) + |Y^+| + |I^+|. \end{aligned}$$

The desired result now follows from the last equality.

Proposition 2. *Suppose that (LP) is feasible and non-degenerate. Let the vector (x^*, y^*, I^*) be a basic optimal solution for (LP) and let the vector $(\lambda^*, \omega^*, \nu^*, v^*)$ be the corresponding optimal solution for (D). Then,*

1. *For each $(j,t) \notin B$, $x^*_{ijt} = 1$ if and only if*

$$a_{ijt} + \lambda^*_{it} d_{jt} + \nu^*_{it} d_{jt} = \min_{k=1,\ldots,m} (a_{kjt} + \lambda^*_{kt} d_{jt} + \nu^*_{kt} d_{jt})$$

and

$$a_{ijt} + \lambda^*_{it} d_{jt} + \nu^*_{it} d_{jt} < \min_{k=1,\ldots,m;\, k \neq i} (a_{kjt} + \lambda^*_{kt} d_{jt} + \nu^*_{kt} d_{jt}).$$

2. *For each $(j,t) \in B$, there exists an index i such that*

$$a_{ijt} + \lambda^*_{it} d_{jt} + \nu^*_{it} d_{jt} = \min_{k=1,\ldots,m;\, k \neq i} (a_{kjt} + \lambda^*_{kt} d_{jt} + \nu^*_{kt} d_{jt}).$$

Proof. Observe that

$$\lambda^*_{it} = \min_{l=1,\ldots,q} (c_{lit} + \omega_{lt}) \geq 0$$

for $j = 1, \ldots, n$; $t = 1, \ldots, T$, and by using the nonnegativity of vector λ^* we have that

$$v^*_{jt} = \min_{i=1,\ldots,m} (a_{ijt} + \lambda^*_{it} d_{jt} + \nu^*_{it} d_{jt}) \geq 0$$

for $j = 1, \ldots, n$; $t = 1, \ldots, T$.

Thus, without loss of optimality, we can add to (D) the nonnegativity constraints on the vectors λ and v. By adding slack variables s_{ijt}, S_{lit} and U_{it} to the constraints in (D), we can reformulate it as

$$\text{maximize} \sum_{t=1}^{T}\sum_{j=1}^{n} v_{jt} - \sum_{t=1}^{T}\sum_{l=1}^{q} b_{lt}\omega_{lt} - \sum_{t=1}^{T}\sum_{i=1}^{m} r_{it}\nu_{it}$$

subject to (D′)

$$\begin{aligned}
v_{jt} + s_{ijt} &= a_{ijt} + \lambda_{it}d_{jt} + \nu_{it}d_{jt} \\
&\qquad i = 1,\ldots,m;\ j = 1,\ldots,n;\ t = 1,\ldots,T \\
\lambda_{it} + S_{lit} &= \omega_{lt} + c_{lit} \\
&\qquad i = 1,\ldots,m;\ l = 1,\ldots,q;\ t = 1,\ldots,T \\
-\lambda_{it} + \lambda_{i,t+1} + U_{it} &= h_{it} \quad i = 1,\ldots,m;\ t = 1,\ldots,T-1 \\
\lambda_{it} &\geq 0 \quad i = 1,\ldots,m;\ t = 1,\ldots,T \\
\omega_{lt} &\geq 0 \quad l = 1,\ldots,q;\ t = 1,\ldots,T \\
\nu_{it} &\geq 0 \quad i = 1,\ldots,m;\ t = 1,\ldots,T \\
v_{jt} &\geq 0 \quad j = 1,\ldots,n;\ t = 1,\ldots,T \\
s_{ijt} &\geq 0 \quad i = 1,\ldots,m;\ j = 1,\ldots,n;\ t = 1,\ldots,T \\
S_{lit} &\geq 0 \quad l = 1,\ldots,q;\ i = 1,\ldots,m;\ t = 1,\ldots,T \\
U_{it} &\geq 0 \quad i = 1,\ldots,m;\ t = 1,\ldots,T-1.
\end{aligned}$$

Let $(\lambda^*, \omega^*, \nu^*, v^*, s^*, S^*, U^*)$ be the optimal solution for (D′). For each $(j,t) \in B$, there exist at least two variables x^*_{ijt} that are strictly positive. Hence, by the complementary slackness conditions, there exist at least two variables s^*_{ijt} equal to zero. This proves Claim 2.

To prove Claim 1, it is enough to show that for each $(j,t) \notin B$ there exists exactly one variable $s^*_{ijt} = 0$. By complementary slackness conditions we know that at least there exists one of these variables. We have to show the uniqueness, and we do it by counting the variables at level zero in the vector $(\lambda^*, \omega^*, \nu^*, v^*, s^*, S^*, U^*)$. There are at least $qT - |Q|$ variables ω^*_{lt}, $mT - |W|$ variables ν^*_{it}, $|F|$ variables s^*_{ijt} corresponding to $(j,t) \in B$, $nT - |B|$ variables s^*_{ijt} corresponding to $(j,t) \notin B$, $|Y^+|$ variables S^*_{lit} equal to zero, and $|Y^+|$ variables U^*_{it} equal to zero. In total, we have at least $qT - |Q| + mT - |W| + |F| + nT - |B| + |Y^+| + |I^+| = qT + 2mT + nT$ zeroes in the optimal dual solution, where the last equality follows from Lemma A.1. So, these are exactly all the variables at level zero in vector $(\lambda^*, \omega^*, \nu^*, v^*, s^*, S^*, U^*)$. Then, for each $(j,t) \notin B$ there exists exactly one variable $s^*_{ijt} = 0$, and Claim 1 follows.

Lemma 3. *Let $\delta \in \mathbb{R}^{mT}_+$ and $\gamma \in \mathbb{R}^{qT}_+$. Then there exists a vector $y \in \mathbb{R}^{qmT}_+$ such that*

$$\sum_{i=1}^{m} y_{lit} \leq \gamma_{lt} \qquad \text{for all } l = 1,\ldots,q;\ t = 1,\ldots,T \tag{10}$$

and

$$\sum_{\tau=1}^{t}\sum_{l=1}^{q} y_{li\tau} \geq \sum_{\tau=1}^{t} \delta_{i\tau} \qquad \text{for all } i = 1,\ldots,m;\ t = 1,\ldots,T \tag{11}$$

if and only if

$$\sum_{\tau=1}^{t}\sum_{i=1}^{m} \delta_{i\tau} \leq \sum_{\tau=1}^{t}\sum_{l=1}^{q} \gamma_{l\tau} \qquad \text{for all } t = 1,\ldots,T. \tag{12}$$

Proof. It can easily be seen that condition (12) is necessary to ensure the existence of a nonnegative vector $y \in \mathbb{R}_+^{qmT}$ satisfying (10) and (11). We will show, by induction on t, that this condition is also sufficient.

For $t = 1$, the inequalities in conditions (10) and (11) together with the nonnegativity assumption on y define the feasible region of a standard transportation problem. Moreover, the inequality in condition (12) for $t = 1$ says that the aggregate demand cannot exceed the aggregate capacity, which clearly is a sufficient condition for feasibility of the standard transportation problem.

Now, we will assume that if the inequality conditions in (12) hold for $t = 1,\ldots,t'$, then there exists a nonnegative vector $y \in \mathbb{R}_+^{qmt'}$ so that the inequalities in conditions (10) and (11) are satisfied for $t = 1,\ldots,t'$. We will show that the same result holds for horizon $t' + 1$.

We will distinguish two cases, depending on the difference between the aggregate demand in period $t' + 1$ and the aggregate capacity in the same period. First consider the case where the aggregate demand is no more than the aggregate capacity, i.e.,

$$\sum_{i=1}^{m} \delta_{i,t'+1} \leq \sum_{l=1}^{q} \gamma_{l,t'+1}.$$

Then there exists a vector $z \in \mathbb{R}_+^{qm}$ such that

$$\sum_{i=1}^{m} z_{li,t'+1} \leq \gamma_{l,t'+1} \qquad l = 1,\ldots,q$$

and

$$\sum_{l=1}^{q} z_{li,t'+1} \geq \delta_{i,t'+1} \qquad i = 1,\ldots,m.$$

Moreover, by the induction hypothesis there exists a vector $y \in \mathbb{R}_+^{qmt'}$ such that

$$\sum_{i=1}^{m} y_{lit} \leq \gamma_{lt} \qquad l = 1,\ldots,q;\ t = 1,\ldots,t' \tag{18}$$

and

$$\sum_{\tau=1}^{t}\sum_{l=1}^{q} y_{li\tau} \geq \sum_{\tau=1}^{t} \delta_{i\tau} \qquad i = 1, \ldots, m;\ t = 1, \ldots, t'. \tag{19}$$

It is easy to see that (y, z) a nonnegative vector satisfying the inequalities in conditions (10) and (11) for $t = 1, \ldots, t' + 1$.

Next, we will consider the case where

$$\sum_{i=1}^{m} \delta_{i,t'+1} > \sum_{l=1}^{q} \gamma_{l,t'+1}.$$

It suffices to show that the excess demand in period $t' + 1$, i.e.,

$$\sum_{i=1}^{m} \delta_{i,t'+1} - \sum_{l=1}^{q} \gamma_{l,t'+1}$$

can be supplied in previous periods. This is easy to see since

$$\sum_{\tau=1}^{t}\sum_{i=1}^{m} \delta_{i\tau} \leq \sum_{\tau=1}^{t}\sum_{l=1}^{q} \gamma_{l\tau}$$

for all $t = 1, \ldots, t' - 1$, and

$$\sum_{\tau=1}^{t'}\sum_{i=1}^{m} \delta_{i\tau} + \sum_{i=1}^{m} \delta_{i,t'+1} - \sum_{l=1}^{q} \gamma_{l,t'+1} \leq \sum_{\tau=1}^{t'}\sum_{l=1}^{q} \gamma_{l\tau}.$$

Combinatorial Optimisation Problems of the Assignment Type and a Partitioning Approach

Andreas Klose[1] and Andreas Drexl[2]

[1] Universität St. Gallen, 9000 St. Gallen, and Institut für Operations Research der Universität Zürich, 8015 Zürich, Switzerland
[2] Christian-Albrechts-Universität zu Kiel, Lehrstuhl für Produktion und Logistik, 24908 Kiel, Germany

Abstract Assignment type problems consist in optimally assigning or allocating a given set of "activities" to a given set of "agents". Optimisation problems of the assignment type have numerous applications in production planning and logistics. A popular approach to solve such problems or to compute lower bounds on the optimal solution value (in case of a minimisation problem) is to employ column generation. By means of considering subsets of "activities" which can feasibly be assigned to a single agent, the problem is reformulated as some kind of set-partitioning problem. Column generation is then used in order to solve the linear relaxation of the reformulation. The lower bound obtainable from this approach may, however, be improved by partitioning the set of agents into subsets and by considering subsets of activities which can feasibly be assigned to subsets of agents. This paper outlines the application of this partitioning method to a number of important combinatorial optimisation problems of the assignment type.

1 Introduction

An optimisation problem of the assignment type is to find a feasible least-cost assignment of a given set I of "activities" i to a given set J of "agents" j. Since a seminal work of Koopmans and Beckmann (1957) on linear and quadratic assignment problems, optimisation problems of the assignment type play an important role in Economics and Operations Research. Assignment type problems have numerous applications in logistics and production planning, and many mathematical models of decision problems in logistics contain an assignment type problem as a subproblem.

Formally, an assignment or allocation of activities $i \in I$ to agents $j \in J$ is a mapping $x : I \times J \longrightarrow [0,1]^{|I|\times|J|}$. An assignment problem then consists in determining a feasible assignment $x \in X$ which minimises a given objective function $g(x)$. Assuming that the objective function $g(x)$ as well as the set X of feasible assignments x is decomposable, that is $g(x) = \sum_{j\in J} g_j(x)$ and $X = \left\{x = (x_1, \ldots, x_{|J|}) : x_j \in X_j \subseteq [0,1]^{|I|} \ \forall\, j \in J\right\}$, the problem can be

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

formulated as follows:

$$\min \sum_{j \in J} g_j(x_j) \tag{1a}$$

$$\text{s.t.:} \sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \tag{1b}$$

$$x_j \in X_j \subseteq [0,1]^{|I|} \quad \forall j \in J\,. \tag{1c}$$

In addition, it is assumed that either X is a nonempty finite set or that each X_j is a polyhedron and that each $g_j(x_j)$ is concave.

Assignment problems often involve indivisibilities. Indivisibilities occur

- if an activity has to be assigned to exactly one agent and/or
- a fixed-charge is imposed on the use of an agent.

Indivisibilities often lead to strong $\mathcal{NP}$-hard combinatorial optimisation problems. In order to solve such problems to optimality, the computation of sharp lower bounds on the optimal solution value is required. One possible way to accomplish this, is to reformulate the problem. A popular reformulation is to consider all subsets of activities feasibly assignable to each single agent j. The problem then consists in choosing subsets of activities in such a way that each activity is covered by one of the selected subsets and total costs are minimised. The lower bound resulting from the linear relaxation of this reformulation may, however, be improved by means of partitioning the set of agents into (small) subsets and considering subsets of activities feasibly assignable to subsets of agents. In this paper, this partitioning approach is outlined for some $\mathcal{NP}$-hard optimisation problems of the assignment type. The next section describes some important $\mathcal{NP}$-hard assignment problems. Sect. 3 discusses the standard way of applying column generation to these problems. Sect. 4 introduces the partitioning approach and outlines column generation procedures based on this approach for the generalised assignment problem (GAP) and the capacitated facility location problem (CFLP). Furthermore, some computational results obtained for the CFLP are given. Finally, the findings and some directions for future research are summarised in Sect. 5.

2 Optimisation Problems of the Assignment Type

Formulation (1) of a generic assignment problem excludes optimisation problems with interactions between assigned objects like the quadratic assignment problem. Nevertheless, formulation (1) still covers a large number of assignment type problems ranging from polynomial solvable cases like the matching and transportation problem to strong $\mathcal{NP}$-hard optimisation problems like the generalised assignment problem (GAP), bin-packing problem (BPP), fixed-charge transportation problem (FCTP), and discrete location

problems. Furthermore, many combinatorial assignment problems are difficult in the sense that no polynomial-time approximation scheme (PTAS) exists, unless the class $\mathcal{P}$ of decision problems solvable in polynomial time equals the set $\mathcal{NP}$ of decision problems solvable in nondeterministic polynomial time. Optimisation problems of this type are called MAX-$\mathcal{SNP}$-hard. An example of a MAX-$\mathcal{SNP}$-hard assignment type problem is the uncapacitated facility location problem (see Arora and Lund (1997) for a proof). Moreover, some assignment type problems are $\mathcal{APX}$-complete, that is the existence of a PTAS for such an optimisation problem implies the existence of a PTAS for any other $\mathcal{NP}$-hard optimisation problem. An example of an $\mathcal{APX}$-complete optimisation problem is the GAP. In the case of the GAP, even the question if a feasible solution exists, is an $\mathcal{NP}$-complete problem. For an overview on complexity theory of optimisation problems see Crescenzi and Kann (1998) and Hochbaum (1997). The following discussion is restricted to "difficult" assignment problems. Some of these problems and areas of applications are described below.

2.1 Generalised Assignment Problem

The GAP is to optimally assign a set $I = \{1, \ldots, m\}$ of tasks to a set $J = \{1, \ldots, n\}$ of agents. Mathematically, the problem can be formulated as follows

$$\min \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \tag{2a}$$

$$\text{s.t.:} \sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \tag{2b}$$

$$\sum_{i \in I} d_{ij} x_{ij} \leq s_j \quad \forall j \in J \tag{2c}$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in I, j \in J, \tag{2d}$$

where c_{ij} is the cost of assigning task i to agent j, s_j is agent's j capacity, and d_{ij} denotes the amount of resources required by agent j to perform task i. The binary variable x_{ij} is equal to 1 if agent j performs task i and 0 otherwise. In case that processing tasks by agents requires more than one type of resource, the problem is known as the multi-resource GAP.

A large number of solution methods for the GAP have been proposed in the literature. Fisher et al. (1986) and Guignard and Rosenwein (1989) use Lagrangian ascent methods, that is dual-based procedures employing Lagrangian relaxation of the semi-assignment constraints (2b). Cattrysse et al. (1994) reformulate the GAP as a set-partitioning problem and apply a heuristic procedure for its solution. Savelsbergh (1997) solves the linear relaxation of this reformulation by means of column generation, and proposes a branch-and-price algorithm for computing optimal solutions. Cattrysse et al. (1998)

develop a linear programming heuristic and a branch-and-cut approach based on employing liftet cover inequalitities. Heuristic methods based on local search, tabu search and simulated annealing are described in Osman (1995). For more comprehensive surveys of solution methods for the GAP, we refer to Martello and Toth (1990a), Cattrysse and Van Wassenhove (1992), and Romero Morales (2000).

The GAP is often a subproblem of a larger model of a decision problem in logistics. Examples of such application areas are vehicle routing and distribution system design. Fisher and Jaikumar (1981) describe a "cluster first–route second" vehicle routing heuristic. In a first phase, customer clusters are obtained by means of selecting "seed customers" and solving a GAP in order to feasibly assign the other customers to the selected seeds. Within distribution system design, a GAP has to be solved in order to evaluate selected locations of depots and to optimally assign customer demands to depots if single-sourceing of customers is required. Campell and Langevin (1995) use a mathematical model which is a 2-resource GAP in order to find a low cost assignment of snow removal sectors to snow disposal sites for the City of Montreal. The GAP is a static model of demand allocation. Romero Morales (2000) describes a dynamic model for demand allocation. The model allows time-varying demand patterns and incorporates inventory decisions. The resulting "multi-period single-sourceing problem" is reformulated as a GAP with a convex objective function; greedy heuristics resembling greedy heuristics for the GAP are developed and a branch-and-price approach is proposed.

Another application area of the GAP is tactical and operational planning of flexible manufacturing systems (FMS). A FMS is an automated production system consisting of a set of numerically controlled machines interconnected by an automated transportation system. Each machine is equipped with a tool magazine which can be armed with different tools. Each tool occupies a given number of slots. Tools available at the local tool magazine can be changed by means of an automatic tool changer. Therefore, the FMS can perform different sets of operations and produce different parts in any order, if the FMS is equipped with an appropriate set of tools. Lee and Kim (1998) formulate the order selection problem as a 2-resource GAP. Each order is specified by the due date and the number of parts to be produced. A set of orders to be produced during the planning horizon is given. It has to be decided, which order to produce in which period. Total costs consist of earliness and tardiness costs as well as subcontracting costs. Earliness and tardiness costs are incurred if an order is finished before and after the due date, respectively. Subcontracting costs, however, are incurred if an order is not selected within the planning horizon. Capacity constraints to be taken into account in each period are the total tool magazine capacity as well as total machine processing time capacity. Kuhn (1995) considers the loading problem in FMSs. For a given set of part types that have been selected to be produced simultaneously, the problem is to decide on the assignment of

operations to machines. The assignment of operations determines the assignment of tools to machines. The problem is formulated as an integer program with the objective of minimising the largest workload in such a way that the tool magazine capacity for each machine cannot be violated. In order to solve the problem, Kuhn proposes a heuristic algorithm based on repeatedly solving a GAP. A feasible solution to the GAP assigns operations to machines such that the largest workload of machines in a known feasible solution to the loading problem is reduced. The objective function of the GAP approximates the use of tools slots required by the operations assigned to the machines.

In telecommunication or computer networks, terminals are often connected to the access point of a "backbone network" via so-called concentrators. The terminal layout problem addresses the question of how to interconnect terminals to their associated concentrors. Due to the complexity of the telecommunication network design problem, the concentrator location and the terminal layout problem are usually treated independently. For given locations of the concentrators, the problem of assigning terminals to concentrators can be formulated as a GAP (see e. g. Mirzaian (1985) and Chardaire (1999)). Other applications of the GAP in the area of telecommunication and computer networks concern e. g. the assignment of tasks in a network of functionally similar computers (Balachandran (1976)) or the assignment of user nodes to processing sites with the objective of minimising telecommunication costs subject to capacity constraints on processors (Pirkul (1986)).

2.2 Bin-Packing Problem

Given a set $I = \{1, \ldots, m\}$ of items with weight $w_i > 0$, the bin-packing problem (BPP) is to fit the items into bins of capacity c in such a way that the number of bins used is minimised. Introducing binary variables x_{ij}, which equal 1 if item i is assigned to bin $j \in J$, a mathematical formulation of the problem is

$$\begin{aligned} \min \sum_{j \in J} & g(x_j) \\ \sum_{j \in J} & x_{ij} = 1 \quad \forall i \in I \\ \sum_{i \in I} & w_i x_{ij} \leq c \quad \forall j \in J \\ & x_{ij} \in \{0,1\} \quad \forall i \in I, j \in J, \end{aligned}$$

where $g(x_j) = 1$ if $\sum_{i \in I} x_{ij} > 0$ and 0 otherwise. Defining $y_j = 1$ if $g(x_j) = 1$ and 0 otherwise, the linear formulation

$$\min \sum_{j \in J} y_j \tag{3a}$$

$$\sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \tag{3b}$$

$$\sum_{i \in I} w_i x_{ij} \leq c y_j \quad \forall j \in J \tag{3c}$$

$$x_{ij}, y_j \in \{0,1\} \quad \forall i \in I, j \in J, \tag{3d}$$

is obtained. For the BPP, simple approximation algorithms with a constant, asymptotic worst-case performance ratio are known (see Martello and Toth (1990a)). The linear relaxation of problem (3) can be solved by inspection. The optimal objective function value of the LP-relaxation is given by $r = \sum_{i \in I} w_i / c$. Thus, $\lceil r \rceil$ is a simple lower bound on the mininum number of bins to be used, where $\lceil r \rceil$ is the smallest integer greater or equal than r. Martello and Toth (1990b) improve this bound by partitioning the set I of items into three subsets in such a way that items of the first two subsets require separate bins and that no item of the third subset can be assigned to a bin containing an item from the first subset. Furthermore, they provide a reduction algorithm which checks, if item subsets of cardinality less than 4 have to be packed into the same bin in an optimal solution. Martello and Toth (1990a) describe a branch-and-bound algorithm for the BPP which makes use of this bounding procedure and reduction algorithm.

Heuristic vehicle routing algorithms, as e. g. parallel route building procedures (see e. g. Kontoravdis and Bard (1995)), make use of lower bounds on the minimum number of vehicles required. Such bounds can be obtained by solving bin-packing problems. In this case, the items and their weights are given by the customers and their demands; the bin size is the vehicle capacity. Kontoravdis and Bard (1995) also make use of a maximum route duration constraint in order to obtain bounds on the number of vehicles by means of solving a bin-packing problem. In case of a multiple use of vehicles, the problem of assigning vehicles to tours can also be formulated as a bin-packing problem (see e. g. Fleischmann (1990) and Taillard et al. (1996)).

Another application of the BPP is e. g. the problem of assigning jobs with a given due date to identical machines in such a way that the number of machines used is a minimum and all jobs can be finished before the due date.

2.3 Fixed-Charge Transportation Problem

The fixed-charge transportation problem (FCTP) is obtained from the classical transportation problem by imposing a fixed cost on each transportation link if there is a positive flow on this link. Let $I = \{1, \ldots, m\}$ denote the set of destinations with demands d_i, and let $J = \{1, \ldots, n\}$ denote the set of origins with supplies s_j. The FCTP then consists in solving the mathematical program

$$\min \sum_{i \in I} \sum_{j \in J} g_{ij}(x_{ij}) \tag{4a}$$

$$\text{s.t.:} \sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \tag{4b}$$

$$\sum_{i \in I} d_i x_{ij} \leq s_j \quad \forall j \in J \tag{4c}$$

$$0 \leq x_{ij} \leq u_{ij} \quad \forall i \in I, j \in J, \tag{4d}$$

where x_{ij} is the portion of destination i's demand supplied from origin j, $u_{ij} d_i \leq d_i$ is an upper bound on the amount which can be shipped on link (j,i), and $g_{ij}(x_{ij})$ is the cost of shipping the amount $x_{ij} d_i$ on link (j,i). Assuming that the variable transportation costs are proportional to the amount shipped, the costs $g_{ij}(x_{ij})$ are given by $g_{ij}(x_{ij}) = c_{ij} x_{ij} + f_{ij}$ if $x_{ij} > 0$ and 0 otherwise. By means of introducing binary variables y_{ij} which equal 1 if and only if $x_{ij} > 0$, the problem is easily transformed to the linear mixed-integer program

$$\min \sum_{i \in I} \sum_{j \in J} (c_{ij} x_{ij} + f_{ij} y_{ij}) \tag{5a}$$

$$\text{s.t.:} \sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \tag{5b}$$

$$\sum_{i \in I} d_i x_{ij} \leq s_j \quad \forall j \in J \tag{5c}$$

$$0 \leq x_{ij} \leq u_{ij} y_{ij} \quad \forall i \in I, j \in J \tag{5d}$$

$$y_{ij} \in \{0,1\} \quad \forall i \in I, j \in J. \tag{5e}$$

A number of branch-and-bound algorithms have been proposed to solve the FCTP (see Kennington and Unger (1976), Cabot and Erenguc (1984, 1986), Palekar et al. (1990)). Most of these algorithms make use of penalties in order to fix binary variables and to be able to prematurely prune nodes of the enumeration tree. Since the FCTP is a special case of fixed-charge network flow problems, polyhedral cuts developed for this problem (see e. g. Nemhauser and Wolsey (1988) and Bienstock and Günlük (1996)) may also be used to solve the FCTP. Wright and Haehling von Lanzenauer (1989) develop a Lagrangian heuristic for the FCTP which is based on relaxing the variable upper bound constraints $x_{ij} \leq u_{ij} y_{ij}$. It can be proved that an optimal solution to the FCTP is an extreme point of the convex region defined by constraints (4b)–(4d) (see Hirsch and Dantzig (1968)). Basic feasible solutions to the system (4b)–(4d) define spanning trees of the transportation network. Sun et al. (1998) propose a tabu search procedure for the FCTP which replaces a single link of the current spanning tree by a link not contained in the tree such that a new basic feasible solution is obtained. Göthe-Lundgren and Larsson (1994) consider the pure FCTP. In this case the objective function coefficients c_{ij} of the continuous variables x_{ij} are all equal to zero. The Benders' reformulation of the pure FCTP consists, therefore, only of feasibility cuts. A feasibility cut excludes an integer solution y which does not allow

a feasible flow from the set of origins to the set of destinations. The large set of all feasibility cuts has the structure of a set-covering problem. Thus, by means of applying Benders' reformulation principle, the FCTP is transformed to a set-covering problem. In order to solve the reformulated problem to ϵ-optimality, Göthe-Lundgren and Larsson (1994) apply Benders' decomposition, where violated feasibility cuts are determined by means of solving a maximum flow problem. Furthermore, Göthe-Lundgren and Larsson (1994) propose a Lagrangian relaxation approach which relaxes a large number of the feasibility cuts. In this way, a lower bound as well as a feasible solution to the pure FCTP is computed. Hultberg and Cardoso (1997) describe the "teacher assignment problem", a pure FCTP in which all fixed charges f_{ij} are equal to one. Thus, the problem is to find the most degenerate basic feasible solution to constraints (4b)–(4d). Hultberg and Cardoso (1997) show that this problem is equivalent to a maximum cardinality partition problem and provide a branch-and-bound procedure for computing optimal solutions. Herer et al. (1996) study the FCTP with a single destination. They develop two simple greedy heuristics as well as an implicit enumeration scheme which also makes use of domination rules and lower bounds for accelerating the search.

Herer et al. (1996) present three applications of the FCTP. A first application is to select suppliers and to determine periodic shipment quantities for an item to be procured in such a way that total periodic costs are minimised, periodic demand is met and a supplier's capacity is not exceeded. The total costs consist of purchasing, fixed ordering, inventory carrying and fixed supplier management costs. Periodic inventory costs (ordering and inventory carrying) are computed based on Economic Order Quantity logic. As a second application of the (single-sink) FCTP, Herer et al. (1996) mention the selection of trucks for meeting a firm's delivery needs such that the sum of variable transportation and fixed vehicle costs is as low as possible. As a final application they mention process selection. A pre-specified amount of a number of products can be made using several different processes, each of which has a given capacity and fixed set-up cost. The problem is then to determine which processes to use to what extend so as to minimise costs. Moore et al. (1991) describe an integer transportation model which closely resembles the FCTP and which is used by the central dispatch control center of a metal company in the U.S. for assigning shipments to carriers. An extended version of this model as well as a simulation study were also used for the purposes of core carrier selection.

2.4 Discrete Location Problems

Discrete location problems form a large subfamily of assignment type optimisation problems. Applications of discrete location problems include location and distribution planning (see e. g. Geoffrion and Graves (1974), Gelders et al. (1987), Tüshaus and Wittmann (1998), Engeler et al. (1999), Bruns

et al. (2000)), lotsizing in production planning (Pochet and Wolsey (1988)), telecommunication and computer network design (Mirzaian (1985), Boffey (1989), Chardaire (1999)), vendor selection (Current and Weber (1994)), and physical database design (Caprara and Salazar (1999)). For comprehensive surveys of discrete location problems, we refer the reader to Aikens (1985), Mirchandani and Francis (1990), Daskin (1995), Revelle and Laporte (1996), Domschke and Krispin (1997) and Owen and Daskin (1998).

As one example of a discrete location problem, we describe the capacitated facility location problem (CFLP). It consists in deciding which depots to open from a given set $J = \{1, \ldots, n\}$ of potential depot locations and how to assign a set $I = \{1, \ldots, m\}$ of customers with given demands d_i to those depots. The objective is to minimise total fixed and shipping costs. Constraints are that each customer's demand must be satisfied and that each depot j cannot supply more than its capacity s_j if it is open. Let f_j denote the fixed cost of operating facility j and let c_{ij} denote the cost of supplying all of customer i's demand from location j. The CFLP can then be stated as follows:

$$\begin{aligned}
\min \quad & \sum_{j \in J} g_j(x_j) \\
\text{s.t.:} \quad & \sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \\
& \sum_{i \in I} d_i x_{ij} \leq s_j \quad \forall j \in J \\
& x_{ij} \geq 0 \quad \forall i \in I, j \in J,
\end{aligned}$$

where x_{ij} is the fraction of customer i's demand met from facility j, and $g_j(x_j) = f_j + \sum_{i \in I} c_{ij} x_{ij}$ if $\sum_{i \in I} x_{ij} > 0$ and 0 otherwise. A linear formulation is obtained by introducing binary variables y_j indicating if a facility j is open or not:

$$\min \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{j \in J} f_j y_j \tag{6a}$$

$$\text{s.t.:} \sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \tag{6b}$$

$$\sum_{i \in I} d_i x_{ij} \leq s_j y_j \quad \forall j \in J \tag{6c}$$

$$\sum_{j \in J} s_j y_j \geq \sum_{i \in I} d_i \tag{6d}$$

$$x_{ij} - y_j \leq 0 \quad \forall i \in I, j \in J \tag{6e}$$

$$x_{ij} \geq 0 \quad \forall i \in I, j \in J \tag{6f}$$

$$y_j \in \{0, 1\} \quad \forall j \in J. \tag{6g}$$

In the above formulation, constraints (6e) and (6d) are redundant; however, these constraints help to strengthen certain relaxations of the CFLP.

Numerous heuristic and exact algorithms for the CFLP have been proposed in the literature. These methods include greedy heuristics (Khumawala (1974), Jacobsen (1983), Korupolu et al. (1998)), linear programming based rounding and filtering techniques (Shmoys et al. (1997)), Benders' decomposition (Wentges (1996)), branch-and-cut methods based on polyhedral cuts (Aardal et al. (1995), Aardal (1998)), and a number of Lagrangian relaxation approaches used in heuristics or exact branch-and-bound schemes (Nauss (1978), Christofides and Beasley (1983), Beasley (1988), Ryu and Guignard (1992)).

3 Column Generation Applied to Assignment Problems

Consider the generic assignment problem (1) and assume that the set $X = X_1 \times X_2 \times \cdots \times X_{|J|}$ and thereby each X_j is finite, that is $X_j = \{x_j^t : t \in T_j\}$. The problem may then be rewritten as

$$\min \sum_{j \in J} \sum_{t \in T_j} g_j(x_j^t)\alpha_{jt} \tag{7a}$$

$$\text{s.t.:} \sum_{t \in T_j} \alpha_{jt} = 1 \quad \forall j \in J \tag{7b}$$

$$\sum_{j \in J} \sum_{t \in T_j} x_{ij}^t \alpha_{jt} = 1 \quad \forall i \in I \tag{7c}$$

$$\alpha_{jt} \in \{0,1\} \quad \forall j \in J, t \in T_j\,, \tag{7d}$$

where α_{jt} equals 1 if column x_j^t is chosen for agent j. The reformulation (7) is a set-partioning problem if the columns x_j^t are binary. A lower bound on the optimal solution value of problem (1) can be obtained by relaxing the integrality requirements (7d) and solving the resulting linear program. In case, that each function g_j is convex, it is straightforward to show that this bound is at least as strong as a lower bound obtained by replacing in (1) the set X by a convex set $\overline{X}$ containing X (see e. g. Romero Morales (2000) for a proof).

The number of variables in the reformulation (7) grows exponentially with problem size. Thus, column generation has to be applied in order to solve its linear relaxation. The linear programming dual of (7) is given by

$$\begin{aligned} &\max \sum_{j \in J} \nu_j + \sum_{i \in I} \eta_i \\ &\text{s.t.:} \ \nu_j + \sum_{i \in I} x_{ij}^t \eta_i \leq g_j(x_j^t) \quad \forall j \in J, t \in T_j\,. \end{aligned} \tag{8}$$

For each $j \in J$ let $\{x_j^t : t \in \overline{T}_j\}$ denote known subsets of columns $t \in \overline{T}_j \subseteq T_j$. The restricted linear master problem is obtained from the linear relaxation

of (7) by replacing each T_j with $\overline{T}_j$. Assume that the restricted linear master is feasible (otherwise put e.g. an artificial box around the dual variables) and let $\overline{\alpha}$ and $(\overline{\nu}, \overline{\eta})$ denote an optimal basic solution of the restricted linear master and its dual. The basic solution $\overline{\alpha}$ is an optimal solution to the linear relaxation of (7) if it is also dual feasible, that is

$$\overline{\nu}_j + \sum_{i \in I} x_{ij}^t \overline{\eta}_i \leq g(x_j^t) \quad \forall\, j \in J,\, t \in T_j\,.$$

Therefore, in order to detect a violated dual constraint and a column $t \in T_j$ for some $j \in J$ with negative reduced costs, the following pricing problem has to be solved for all or at least some $j \in J$:

$$\min \Big\{ g_j(x_j) - \sum_{i \in I} \overline{\eta}_i x_{ij} : x_j \in X_j \Big\}\,. \tag{9}$$

If x_j^t, $t \in T_j$, is an optimal solution to (9) with $g_j(x_j^t) - \sum_i \overline{\eta}_i x_{ij}^t < \overline{\nu}_j$ new columns with negative reduced costs are found. These columns are added to the restricted linear master, and the master is reoptimised. The process continues until no column with negative reduced costs exists.

It is straightforward to see that this approach of computing a lower bound for the assignment problem (1) is equivalent to a Lagrangian relaxation of the semi-assignment constraints (1b). Dualising constraints (1b) with dual variables η_i gives the Lagrangian subproblem

$$\begin{aligned} L_1(\eta) &= \sum_{i \in I} \eta_i + \min \Big\{ \sum_{j \in J} \Big(g_j(x_j) - \sum_{i \in I} \eta_i x_{ij} \Big) : x_j \in X_j \;\forall\, j \in J \Big\} \\ &= \sum_{i \in I} \eta_i + \sum_{j \in J} \min \Big\{ g_j(x_j) - \sum_{i \in I} \eta_i x_{ij} : x_j \in X_j \Big\} \\ &= \sum_{i \in I} \eta_i + \sum_{j \in J} \min_{t \in T_j} \Big\{ g_j(x_j^t) - \sum_{i \in I} \eta_i x_{ij}^t \Big\}\,, \end{aligned} \tag{10}$$

where the last equality follows from the finiteness of the sets X_j. Setting $\nu_j \equiv \min_{t \in T_j} \{ g_j(x_j^t) - \sum_i \eta_i x_{ij}^t \}$, the problem of maximising the lower bound $L_1(\eta)$ leads then to the linear programming dual (8) of the reformulation (7).

The function $L_1(\eta)$ is piecewise linear and concave. A number of methods having their origins in nondifferential optimisation is, therefore, applicable for the purposes of solving the linear relaxation of problem (7). Examples of such methods are mixtures of Dantzig-Wolfe decomposition and subgradient optimisation (Guignard and Zhu (1994), Klose and Drexl (2001)), bundle methods (Lemaréchal (1989)), and interior point methods (Goffin et al. (1992)). All these methods usually show a better convergence behaviour than the standard column generation procedure which is also known as Dantzig-Wolfe decomposition (Dantzig and Wolfe (1960)) or Kelley's cutting plane algorithm (Kelley (1960)). The suitability of a method for maximising the

function $L_1(\eta)$, however, strongly depends on the specific problem and the hardness of the restricted master problem and pricing subproblem, respectively.

Now consider the second case in which X is not finite, but each X_j is a polyhedron and each function $g_j(x_j)$ is concave. For reasons of simplicity assume in addition, that each set X_j is bounded. Let $\{x_j^t : t \in T_j\}$ denote the set of vertices of X_j. Problem (1) may then be rewritten as follows:

$$\begin{aligned}
\min \quad & \sum_{j \in J} g\Big(\sum_{t \in T_j} \alpha_{jt} x_j^t\Big) \\
\text{s.t.:} \quad & \sum_{t \in T_j} \alpha_{jt} = 1 \quad \forall j \in J \\
& \sum_{j \in J} \sum_{t \in T_j} x_{ij}^t \alpha_{jt} = 1 \quad \forall i \in I \\
& \alpha_{jt} \geq 0 \quad \forall j \in J,\, t \in T_j\,.
\end{aligned}$$

Since g_j is concave, we have

$$g_j\Big(\sum_{t \in T_j} \alpha_{jt} x_j^t\Big) \geq \sum_{t \in T_j} \alpha_{jt} g_j(x_j^t)\,.$$

The linear program

$$\begin{aligned}
\min \quad & \sum_{j \in J} \sum_{t \in T_j} \alpha_{jt} g(x_j^t) \\
\text{s.t.:} \quad & \sum_{t \in T_j} \alpha_{jt} = 1 \quad \forall j \in J \\
& \sum_{j \in J} \sum_{t \in T_j} x_{ij}^t \alpha_{jt} = 1 \quad \forall i \in I \\
& \alpha_{jt} \geq 0 \quad \forall j \in J,\, t \in T_j
\end{aligned}$$

is, therefore, a relaxation of the original problem and its optimal objective function value is a lower bound on the optimal solution value of (1). In order to solve the above relaxation, column generation can be applied in the same way as already described. Obviously, the same relaxation results if the semi-assignment constraints (1b) are relaxed in a Lagrangian manner; due to the concavity of the functions g_j, optimal solutions to the Lagrangian subproblem are vertices x_j^t of the set X_j.

The following examples further illustrate the principle of the reformulation (7).

Example 1. Consider the GAP (2). For each agent $j \in J$ let $\{x_j^t : t \in T_j\}$ denote the set of feasible assignments of tasks $i \in I$ to agent j, that is

$$\{x_j^t : t \in T_j\} = \Big\{x_j \in \{0,1\}^{|I|} : \sum_{i \in I} d_{ij} x_{ij} \leq s_j\Big\}\,.$$

Furthermore, let $C_{jt} = \sum_{i \in I} c_{ij} x_{ij}^t$ be the cost of such an assignment. The GAP may then be formulated as the set-partitioning problem

$$\min \sum_{j \in J} \sum_{t \in T_j} C_{jt} \alpha_{jt} \tag{11a}$$

$$\text{s.t.:} \sum_{t \in T_j} \alpha_{jt} = 1 \quad \forall j \in J \tag{11b}$$

$$\sum_{j \in J} \sum_{t \in T_j} x_{ij}^t \alpha_{jt} = 1 \quad \forall i \in I \tag{11c}$$

$$\alpha_{jt} \in \{0,1\} \quad \forall j \in J, t \in T_j \,. \tag{11d}$$

For each $j \in J$ the pricing subproblem is given by the binary knapsack problem

$$\min \Big\{ \sum_{i \in I} (c_{ij} - \overline{\eta}_i) x_{ij} : \sum_{i \in I} d_{ij} x_{ij} \leq s_j,\ x_{ij} \in \{0,1\}\ \forall i \in I \Big\}, \tag{12}$$

where $(\overline{\nu}, \overline{\eta})$ is an optimal dual solution to the restricted linear master problem. New columns x_j^t price out if $\overline{\nu}_j > \sum_{i \in I} (c_{ij} - \overline{\eta}_i) x_{ij}^t$.

Example 2. Consider the BPP (3). Let $I_j^t \subseteq I$ be a nonempty subset of items which fit into bin j, and let $\{I_j^t : t \in T_j\}$ denote the set of all such subsets. Since all bins have the same capacity c, we have $T_j = T$ and $I_j^t = I^t$ for all $j \in J$. Furthermore, the evaluation $g(I_j^t) = 1$ of a nonempty subset I_j^t does not depend on j. The BPP may, therefore, be rewritten as follows:

$$\begin{aligned}
&\min \sum_{t \in T} \alpha_t \\
&\text{s.t.:} \sum_{t \in T} \alpha_t \leq n \\
&\sum_{t \in T} x_i^t \alpha_t = 1 \quad \forall i \in I \\
&\alpha_t \in \{0,1\} \quad \forall t \in T \,,
\end{aligned}$$

where $x_i^t = 1$ if $i \in I^t$ and 0 otherwise. Since it is assumed that the BPP (3) has a feasible solution, the constraint $\sum_{t \in T} \alpha_t \leq n$ can be dropped and the BPP is reformulated as a pure set-partitioning problem. The pricing subproblem is to solve the single binary knapsack problem

$$\max \Big\{ \sum_{i \in I} \overline{\eta}_i x_i : \sum_{i \in I} w_i \leq c,\ x_i \in \{0,1\}\ \forall i \in I \Big\},$$

where $\overline{\eta}$ is an optimal dual solution to the restricted linear master problem. A new column x^t prices out if $1 - \sum_{i \in I} \overline{\eta}_i x_i^t < 0$.

Example 3. For the FCTP (5) let $\{(y_j^t, x_j^t) : t \in T_j\}$ denote the set of all link selections $y_j \in \{0,1\}^{|I|}$ and transportation flows $x_j \in [0,1]^{|I|}$ respecting the link capacities u_{ij} as well as source node j's capacity s_j. Exactly speaking, $\{(y_j^t, x_j^t) : t \in T_j\}$ is the vertex set of the convex hull of the set of solutions satisfying constraints

$$\sum_{i \in I} d_i x_{ij} \leq s_j \,, 0 \leq x_{ij} \leq u_{ij} y_{ij} \; \forall\, i \in I \text{ and } y_{ij} \in \{0,1\} \; \forall\, i \in I \,.$$

This leads to the reformulation

$$\begin{aligned}
\min \; & \sum_{j \in J} \sum_{t \in T_j} C_{jt} \alpha_{jt} \\
\text{s.t.:} \; & \sum_{t \in T_j} \alpha_{jt} = 1 \quad \forall\, j \in J \\
& \sum_{j \in J} \sum_{t \in T_j} x_{ij}^t \alpha_{jt} = 1 \quad \forall\, i \in I \\
& \alpha_{jt} \in \{0,1\} \quad \forall\, j \in J, t \in T_j \,,
\end{aligned}$$

where $C_{jt} = \sum_{i \in I} (c_{ij} x_{ij}^t + f_{ij} y_{ij}^t)$. The pricing subproblem consists in solving for each $j \in J$ the program

$$\begin{aligned}
\min \; & \sum_{i \in I} \big((c_{ij} - \overline{\eta}_i) x_{ij} + f_{ij} y_{ij} \big) && \text{(13a)} \\
\text{s.t.:} \; & \sum_{i \in I} d_i x_{ij} \leq s_j && \text{(13b)} \\
& 0 \leq x_{ij} \leq u_{ij} y_{ij} \quad \forall\, i \in I && \text{(13c)} \\
& y_{ij} \in \{0,1\} \quad \forall\, i \in I \,, && \text{(13d)}
\end{aligned}$$

where $(\overline{\nu}, \overline{\eta})$ is an optimal dual solution to the restricted linear master problem. If a slack variable is added to the capacity constraint (13b) and the roles of the source node j and the sink nodes $i \in I$ is reversed, the program (13) is easily recognized as a single-sink FCTP or single-node capacitated flow problem. New columns (y_j^t, x_j^t) price out if $\overline{\nu}_j > \sum_{i \in I} \big((c_{ij} - \overline{\eta}_i) x_{ij}^t + f_{ij} y_{ij}^t \big)$.

Example 4. In case of the CFLP, assume that $\{y^t : t \in T^y\}$ is the set of all depot selections which have enough capacity to meet total demand, that is

$$\{y^t : t \in T^y\} = \Big\{ y \in \{0,1\}^{|J|} : \sum_{j \in J} s_j y_j \geq \sum_{i \in I} d_i \Big\} \,.$$

Furthermore, let $\{x_j^t : t \in T_j^x\}$ denote the vertex set of the set of all feasible flows from depot j to the customers, that is $\sum_{i \in I} d_i x_{ij}^t \leq s_j$ and $x_j^t \in [0,1]^{|I|}$.

The CFLP may then be rewritten as the linear mixed-integer program:

$$\min \sum_{t\in T^y} F_t\alpha_t + \sum_{j\in J}\sum_{t\in T_j^x} C_{jt}\beta_{jt} \tag{14a}$$

$$\text{s.t.:} \sum_{t\in T^y} \alpha_t = 1 \tag{14b}$$

$$\sum_{t\in T^y} y_j^t\alpha_t - \sum_{t\in T_j^x} \beta_{jt} \geq 0 \quad \forall j \in J \tag{14c}$$

$$\sum_{j\in J}\sum_{t\in T_j^x} x_{ij}^t\beta_{jt} = 1 \quad \forall i \in I \tag{14d}$$

$$\alpha_t \in \{0,1\} \quad \forall t \in T^y \tag{14e}$$

$$\beta_{jt} \geq 0 \quad \forall j \in J,\, t \in T_j^x\,, \tag{14f}$$

where $F_t = \sum_{j\in J} f_j y_j^t$ and $C_{jt} = \sum_{i\in I} c_{ij}x_{ij}^t$. Constraint (14b) guarantees that exactly one depot set with sufficient capacity is selected; constraints (14c) state that there can be no flow from a closed facility j, and constraints (14d) guarantee that each customer's demand is met. If $(\overline{\zeta}, \overline{\nu}, \overline{\eta})$ is an optimal dual solution to the restricted linear master problem, the pricing problem consists in solving for each $j \in J$ the continuous knapsack problems

$$\max\Big\{\sum_{i\in I}(\overline{\eta}_i - c_{ij})x_{ij} : \sum_{i\in I} d_i x_{ij} \leq s_j,\, 0 \leq x_{ij} \leq 1\ \forall i \in I\Big\}$$

and the binary knapsack problem

$$\min\Big\{\sum_{j\in J}(f_j - \overline{\nu}_j)y_j : \sum_{j\in J} s_j y_j \geq \sum_{i\in I} d_i\,,\ y_j \in \{0,1\}\ \forall j \in J\Big\}\,.$$

New columns x_j^t and y^t price out if $\overline{\nu}_j < \sum_{i\in I}(\overline{\eta}_i - c_{ij})x_{ij}^t$ and $\overline{\zeta} > \sum_{j\in J}(f_j - \overline{\nu}_j)y_j^t$, respectively.

Remark 1. As already shown when discussing the reformulation of the assignment problem (1), in all the aforementioned cases, the linear relaxation of the integer master program can be obtained by relaxing the semi-assignment constraints in a Lagrangian fashion, rewriting the Lagrangian dual as a linear program and dualising this linear program.

4 A Partitioning Approach

Consider the generic assignment problem (1) and assume that each X_j is a finite set (similar arguments apply if each X_j is a polyhedron and each g_j is a concave function). Partition the set J into subsets J_q, $q \in Q$ and $J_l \cap J_q = \emptyset$

for $l \neq q$, in such a way that at least one subset J_q has cardinality greater than one. The semi-assignment constraints (1b) imply the constraints

$$\sum_{j \in J_q} x_{ij} \leq 1 \quad \forall i \in I, q \in Q\,. \tag{15}$$

Thus, if constraints (1b) are relaxed in a Lagrangian manner, the addition of the constraints (15) can help to sharpen the relaxation. Dualising the semi-assignment constraints (1b) with multipliers η_i while adding the redundant constraints (15) gives the Lagrangian subproblem

$$L_2(\eta) \equiv \sum_{i \in I} \eta_i + \min \sum_{q \in Q} \sum_{j \in J_q} \Big\{ g_j(x_j) - \sum_{i \in I} \eta_i x_{ij} \Big\} \tag{16a}$$

$$\text{s.t.:} \sum_{j \in J_q} x_{ij} \leq 1 \quad \forall i \in I, q \in Q \tag{16b}$$

$$x_j \in X_j \quad \forall j \in J\,, \tag{16c}$$

which decomposes into the $|Q|$ subproblems

$$\nu_q = \min \sum_{j \in J_q} \Big(g_j(x_j) - \sum_{i \in I} \eta_i x_{ij} \Big) \tag{17a}$$

$$\text{s.t.:} \sum_{j \in J_q} x_{ij} \leq 1 \quad \forall i \in I \tag{17b}$$

$$x_j \in X_j \quad \forall j \in J_q \tag{17c}$$

such that

$$L_2(\eta) = \sum_{i \in I} \eta_i + \sum_{q \in Q} \nu_q\,.$$

The subproblems (17) have the same structure as the original problem (1) if a dummy agent with assignment costs of zero is added; they are, however, usually much smaller than the original problem. Let $\{(x_j^h)_{j \in J_q} : h \in H_q\}$ denote the set of solutions satisfying constraints (17b) and (17c). The Lagrangian dual, which is the problem of maximising the piecewise linear and concave function $L_2(\eta)$, may then be rewritten as the linear program:

$$\max_{\eta} L_2(\eta) = \max \sum_{i \in I} \eta_i + \sum_{q \in Q} \nu_q$$

$$\text{s.t.:} \; \nu_q + \sum_{i \in I} \Big(\sum_{j \in J_q} x_{ij}^h \Big) \eta_i \leq \sum_{j \in J_q} g_j(x_j^h) \quad \forall q \in Q\,, h \in H_q\,.$$

Dualising this linear program with dual variables μ_{qh} gives:

$$\min \sum_{q\in Q}\sum_{h\in H_q} G_{qh}\mu_{qh} \tag{18a}$$

$$\text{s.t.:} \sum_{h\in H_q} \mu_{qh} = 1 \quad \forall q \in Q \tag{18b}$$

$$\sum_{q\in Q}\sum_{h\in H_q}\Big(\sum_{j\in J_q} x_{ij}^h\Big)\mu_{qh} = 1 \quad \forall i \in I \tag{18c}$$

$$\mu_{qh} \geq 0 \quad \forall q \in Q,\, h \in H_q\,, \tag{18d}$$

where $G_{qh} = \sum_{j\in J_q} g_j(x_j^h)$. The linear program (18) is the linear relaxation of an equivalent reformulation of problem (1). This reformulation is obtained if the nonnegativity constraints (18d) are replaced by $\mu_{qh} \in \{0,1\}\,\forall q,h$. If the x_j^h are binary, $I_q^h = \{i \in I : \sum_{j\in J_q} x_{ij}^h = 1\}$ is a subset of activities which can feasibly be assigned to the subset J_q of agents. The reformulated problem than simply states, that the assignment problem is to select for each subset J_q of agents exactly one subset of activities I_q^h in such a way that each activity is contained in one of the selected subsets and total costs are minimised. The following two examples further illustrate the relaxation (18) and the corresponding reformulation of the assignment problem.

Example 5. Adding constraints (15) to the GAP (2) for a given partitioning $\{J_q : q \in Q\}$ of the agent set J and dualising the semi-assignment constraints (2b) with multipliers η_i gives the Lagrangian subproblem

$$\begin{aligned} L_2(\eta) = \sum_{i\in I}\eta_i + \min \sum_{q\in Q}\sum_{j\in J_q}\sum_{i\in I} c_{ij}x_{ij} \\ \text{s.t.:} \sum_{j\in J_q} x_{ij} \leq 1 \quad \forall i \in I,\, q \in Q \\ \sum_{i\in I} d_{ij}x_{ij} \leq s_j \quad \forall j \in J_q,\, q \in Q \\ x_{ij} \in \{0,1\} \quad \forall i \in I,\, j \in J\,. \end{aligned}$$

The Lagrangian subproblem decomposes into the $|Q|$ subproblems

$$\nu_q = \min \sum_{i\in I}\sum_{j\in J_q} c_{ij}x_{ij} \tag{19a}$$

$$\text{s.t.:} \sum_{j\in J_q} x_{ij} \leq 1 \quad \forall i \in I \tag{19b}$$

$$\sum_{i\in I} d_{ij}x_{ij} \leq s_j \quad \forall j \in J_q \tag{19c}$$

$$x_{ij} \in \{0,1\} \quad \forall i \in I\,, j \in J_q \tag{19d}$$

such that $L_2(\eta) = \sum_{i\in I}\eta_i + \sum_{q\in Q}\nu_q$. Each of the above subproblems is easily recognized as a GAP if a dummy task 0_q with assignment costs $c_{i0_q} = 0\,\forall i$, resource requirements $d_{i0_q} = 1\,\forall i$, and capacity $s_{0_q} = |I|$ is added to the subsets J_q of agents. If $\{(x_j^h)_{j\in J_q} : h \in H_q\}$ denotes the set of feasible solutions of subproblem q, the Lagrangian dual problem $\max_\eta L_2(\eta)$ can be written as the linear program

$$\max \sum_{i\in I}\eta_i + \sum_{q\in Q}\nu_q$$
$$\text{s.t.: } \nu_q + \sum_{i\in I}\Big(\sum_{j\in J_q} x_{ij}^h\Big)\eta_i \le C_{qh} \quad \forall\, q\in Q,\, h\in H_q\,,$$

where $C_{qh} = \sum_{i\in I}\sum_{j\in J_q} x_{ij}^h$. Dualising this program with dual variables μ_{qh} then gives the primal linear master problem

$$\min \sum_{q\in Q}\sum_{h\in H_q} C_{qh}\mu_{qh} \tag{20a}$$
$$\text{s.t.: } \sum_{h\in H_q}\mu_{qh} = 1 \quad \forall\, q\in Q \tag{20b}$$
$$\sum_{q\in Q}\sum_{h\in H_q}\Big(\sum_{j\in J_q} x_{ij}^h\Big)\mu_{qh} = 1 \quad \forall\, i\in I \tag{20c}$$
$$\mu_{qh} \ge 0 \quad \forall\, q\in Q,\, h\in H_q\,, \tag{20d}$$

which is the linear relaxation of a set-partitioning reformulation of the GAP. Columns of this reformulation correspond to subsets of tasks feasibly assignable to subsets J_q of agents.

Example 6. In case of the CFLP (6) the situation is more difficult than for the GAP (2). The aggregate capacity constraint (6d) must be dropped in order to keep the Lagrangian subproblem decomposable, if the demand constraints (6b) are relaxed and the constraints (15) are added. Doing this and dualising constraints (6b) with multipliers η_i the Lagrangian subproblem is

$$L_2(\eta) = \sum_{i\in I}\eta_i + \sum_{q\in Q}\nu_q \tag{21a}$$

where

$$\nu_q = \min \sum_{i\in I}\sum_{j\in J_q} c_{ij}x_{ij} + \sum_{j\in J_q} f_j y_j \tag{21b}$$
$$\text{s.t.: } \sum_{j\in J_q} x_{ij} \le 1 \quad \forall\, i\in I \tag{21c}$$
$$\sum_{i\in I} d_i x_{ij} \le s_j y_j \quad \forall\, j\in J_q \tag{21d}$$
$$0 \le x_{ij} \le y_j \quad \forall\, i\in I,\, j\in J_q \tag{21e}$$
$$y_j \in \{0,1\} \quad \forall\, j\in J_q\,. \tag{21f}$$

The above subproblem can be transformed to a CFLP by adding a dummy depot 0_q with capacity $s_{0_q} = \sum_{i \in I} d_i$ and costs $f_{0_q} = c_{i0_q} = 0 \, \forall i$ to the subset J_q of depots. If $\{(y_j^h, x_j^h)_{j \in J_q}) : h \in H_q\}$ denotes the vertex set of the set of feasible solutions to subproblem q and C_{qh} is defined as

$$C_{qh} = \sum_{i \in I} \sum_{j \in J_q} x_{ij}^h + \sum_{j \in J_q} f_j y_j^h \,,$$

the Lagrangian dual problem and the primal linear master problem can be written in the same way as in the case of the GAP. In this case, the linear primal master problem is the linear relaxation of a program which restates the CFLP as a pure integer program. Columns of this reformulation correspond to feasible flows from subsets J_q of depots to the set I of customers.

Remark 2. In an analogous way, the relaxation (18) can be applied to the BPP (3) and the FCTP (5). In case of the FCTP the resulting Lagrangian subproblem (pricing subproblem) decomposes into smaller FCTPs if a dummy sink is added to each of the $|Q|$ subproblems. In case of the BPP, however, each of the $|Q|$ resulting subproblems is a "mixture" of a BPP and a multiple knapsack problem where there is a profit for each item assigned and a fixed cost of 1 for each knapsack (bin) used.

Remark 3. The discussed examples of assignment type problems include only linear constraints; the requirements $x_j \in X_j$ are capacity constraints of the form $\sum_{i \in I} d_{ij} x_{ij} \leq s_j$. An alternative way of partitioning the problem is, therefore, to decompose the set I of activities into disjoint subsets I_r, $r \in R$, adding the implied constraints $\sum_{i \in I_r} d_{ij} x_{ij} \leq s_j$ for each $j \in J$ and $r \in R$, and relaxing the capacity constraints in a Lagrangian manner. The resulting Lagrangian subproblem decomposes again into smaller subproblems of the same or a similar structure as the original problem. It is, however, not difficult to show that this relaxation is usually not as strong as the relaxation (7) or even as strong as the conventional Lagrangian relaxation of the semi-assignment constraints (1b).

An apparent question which arises, is how to partion the subset J into subsets J_q. Some hints can be derived from the following proposition:

Proposition 1. *Let $\overline{\eta}$ denote an optimal solution to* $\max_\eta L_1(\eta)$, *where the Lagrangian function $L_1(\eta)$ is defined in* (10). *Furthermore, let $\{x^t : t \in \overline{T}\}$ denote the set of optimal solutions to the Lagrangean subproblem* (10) *with $\eta = \overline{\eta}$, that is*

$$L_1(\overline{\eta}) = \sum_{i \in I} \overline{\eta}_i + \sum_{j \in J} \Big(g_j(x_j^t) - \sum_{i \in I} \overline{\eta}_i x_{ij}^t \Big) \quad \forall t \in \overline{T} \,.$$

If $\sum_{j \in J_q} x_{ij}^t \leq 1$ holds for every $q \in Q$ and $t \in \overline{T}$ then $\max_\eta L_1(\eta) = \max_\eta L_2(\eta)$.

Proof. The linear relaxation of (7) and the corresponding dual program (8) can be rewritten in aggregated form as

$$\begin{aligned} \max_{\eta} L_1(\eta) = \min \; & \sum_{t \in T} g(x^t)\alpha_t \\ \text{s.t.:} \; & \sum_{t \in T} \alpha_t = 1 \\ & \sum_{j \in J} \sum_{t \in T} x_{ij}^t \alpha_t = 1 \quad \forall i \in I \\ & \alpha_t \geq 0 \quad \forall t \in T \end{aligned} \tag{22}$$

and

$$\max_{\eta} L_1(\eta) = \max\Big\{ \sum_{i \in I} \eta_i + \nu : \sum_{i \in I} \sum_{j \in J} \eta_i x_{ij}^t + \nu \leq g(x^t) \; \forall t \in T \Big\}, \tag{23}$$

respectively, where $\{x^t : t \in T\} = X = \bigcup_j X_j$ and $g(x^t) = \sum_{j \in J} g_j(x_j^t)$. Analogously, if $\{x^h : h \in H\}$ is the set of solutions $x \in X$ satisfying constraints (15), the aggregated version of the linear program (18) reads:

$$\begin{aligned} \max_{\eta} L_2(\eta) = \min \; & \sum_{h \in H} g(x^h)\mu_h \\ \text{s.t.:} \; & \sum_{h \in H} \mu_h = 1 \\ & \sum_{j \in J} \sum_{h \in H} x_{ij}^h \mu_h = 1 \quad \forall i \in I \\ & \mu_h \geq 0 \quad \forall h \in H\,. \end{aligned} \tag{24}$$

Let $\overline{\alpha}$ and $(\overline{\nu}, \overline{\eta})$ denote an optimal solution to (22) and (23), respectively. From complementary slackness it follows that

$$\overline{\alpha}_t > 0 \;\Rightarrow\; g(x^t) - \sum_{i \in I} \sum_{j \in J} \overline{\eta}_i x_{ij}^t = \overline{\nu} \equiv \min_{t \in T} \Big\{ g(x^t) - \sum_{i \in I} \sum_{j \in J} \overline{\eta}_i x_{ij}^t \Big\}.$$

Therefore, $\{x^t : \overline{\alpha}_t > 0\}$ is a subset of the set $\overline{T}$ of optimal solutions to the Lagrangian subproblem (10) for optimal multipliers $\eta = \overline{\eta}$ (if (22) is not degenerate then $\overline{T} = \{x^t : \overline{\alpha}_t > 0\}$). Since $\sum_{j \in J_q} x_{ij}^t \leq 1$ for all $t \in \overline{T}$ and $q \in Q$, the columns $t \in \overline{T}$ are feasible for (24), that is $\overline{T} \subseteq H$. The solution $\overline{\mu}_t = \overline{\alpha}_t$ if $t \in \overline{T}$ and $\overline{\mu}_t = 0$ if $t \in H \setminus \overline{T}$ is, therefore, a feasible solution to (24) with objective function value $\sum_{h \in H} \overline{\mu}_h x^h = \sum_{t \in \overline{T}} \overline{\alpha}_t x^t = L_1(\overline{\eta}) = \max_\eta L_1(\eta)$. Since generally $L_2(\eta) \geq L_1(\eta)$, the desired result follows.

The above proposition states, that the inequalities (15) must cut-off at least one optimal solution of the Lagrangian subproblem (10) for given optimal Lagrangian multipliers; otherwise the relaxation (18) cannot be stronger than

the linear relaxation of the reformulation (7). A plausible way of determining the subsets J_q is, therefore, to first compute (approximate) optimal multipliers for the "conventional" relaxation (10), and afterwards to construct the subsets J_q in such a way that an optimal solution to the Lagrangian subproblem violates at least one of the constraints (15). How this can be done in detail depends on the specific problem on hand. Other problem-specific topics concern algorithms for computing approximate optimal multipliers for the Lagrangian relaxation (10), the implementation of Lagrangian heuristics, the use of Lagrangian probing methods to reduce problem size, the algorithms used to solve the Lagrangian subproblems (10) and (16), the column generation method used for maximising the function L_2 defined in (16), and finally branching rules in case that the relaxation (18) is used within a branch-and-price framework. In the following, we briefly sketch a possible implementation of the described partitioning approach for solving the GAP and an implemention for the CFLP proposed in Klose and Drexl (2001).

4.1 Outline of a Partitioning Procedure for the GAP

In order to possibly sharpen the linear relaxation of the reformulation (11) of the GAP (2) by means of the partitioning approach, it suffices to group the agents $j \in J$ into pairs, that is to decompose the set J of agents into subsets J_q, $q \in Q$, such that $1 \leq |J_q| \leq 2 \, \forall \, q \in Q$ and $|J_q| = 2$ for at least one $q \in Q$. For each $j \in J$ let $\overline{x}_j$ denote an optimal solution of the pricing subproblem (12) for given optimal dual prices $\overline{\eta}$ of the semi-assignment constraints in the linear relaxation of (11). From proposition 1 it follows that

$$\exists \, i \in I : \overline{x}_{ij_q} + \overline{x}_{ik_q} > 1$$

must hold for at least one set $\overline{x} = (\overline{x}_j)_{j \in J}$ of optimal solutions to the pricing subproblems (12) and at least one pair $J_q = \{k_q, j_q\}$ of agents. This suggests to determine pairs J_q of agents by means of solving the matching problem

$$\begin{aligned} \max \; & \sum_{k \in J} \sum_{\substack{j \in J \\ j > k}} w_{kj} z_{kj} \\ \text{s.t.:} \; & \sum_{\substack{k \in J \\ k < j}} z_{kj} + \sum_{\substack{k \in J \\ k > j}} z_{jk} \leq 1 \quad \forall \, j \in J \\ & z_{kj} \in \{0,1\} \quad \forall \, (k,j) \in J \times J, \, k < j \, , \end{aligned} \tag{25}$$

where $w_{kj} = \sum_{i \in I} \overline{x}_{ik} \cdot \overline{x}_{ij}$. The weight w_{kj} counts how many times the solution $\overline{x}$ violates a constraint $x_{ij} + x_{ik} \leq 1$ if agent j and k are matched. In case of a duality gap, optimal solutions $\overline{x} = (\overline{x}_j)_{j \in J}$ of the pricing subproblems (12) for given optimal dual prices $\overline{\eta}$ are not unique. It might, therefore, be useful to match agents j and k even if $w_{kj} = 0$ for a specific solution $\overline{x}$, and a weightening function, as e. g. $w'_{kj} = \sum_{i \in I} (2\overline{x}_{ij}\overline{x}_{ik} + \overline{x}_{ij} + \overline{x}_{ik})$ might

be more appropriate for the purposes of determining agent pairs. A possible implementation of the partitioning approach for the GAP is then to perform the following steps:

1. Apply "conventional" Lagrangian relaxation of the semi-assignment constraints (2b). Compute approximate optimal multipliers $\overline{\eta}$ and a feasible solution to the GAP by means of the multiplier adjustment method (possibly followed by subgradient optimisation) in conjunction with Lagrangian heuristics (see Fisher et al. (1986), Guignard and Rosenwein (1989), Karabakal et al. (1992)).
2. Apply Lagrangian probing techniques in order to fix binary variables x_{ij} to 0 and 1 without loss of optimality (see Guignard et al. (1997)).
3. Decompose the set of agents into pairs by means of solving the matching problem (25).
4. Solve the linear primal master problem (20) by means of a stabilised column generation method. During the column generation use Lagrangian heuristics after each call to the Lagrangian subproblem (pricing subproblem) (19) in order to obtain (improved) feasible solutions to the GAP. Reapply Lagrangian probing if an improved feasible solution has been found.
5. Let μ^* denote the computed optimal solution of the linear program (20) and let x^*, where $x^*_{ij} = \sum_{h \in H_q} \mu^*_{qh} x^h_{ij}$ for $j \in J_q$, denote the corresponding (fractional) solution in terms of the orginal variables. Apply a rounding procedure and/or solve the restricted integer master problem to optimality in order to obtain an improved feasible solution to the GAP.
6. In case that a duality gap remains, use branch-and-price for computing an optimal solution of the GAP. Possible branching rules are $x_{ij} = 0$ vs. $x_{ij} = 1$ for some fractional $\overline{x}_{ij}$, or $\sum_{j \in J_q} x_{ij} = 0$ vs. $\sum_{j \in J_q} x_{ij} = 1$ for some fractional $\sum_{j \in J_q} \overline{x}_{ij}$, or even a multi-branching which generates $|Q|$ branches and forces $\sum_{j \in J_q} x_{ij}$ to one on each of these branches.

4.2 Outline of a Partitioning Procedure for the CFLP

Klose and Drexl (2001) propose the following implementation of the partitioning approach for the CFLP:

1. Apply Lagrangian relaxation of the demand constraints (6b) in formulation (6) of the CFLP. Compute approximate optimal multipliers by means of subgradient optimisation and obtain a feasible solution to the CFLP by means of Lagrangian heuristics.
2. Use Lagrangian probing in order to reduce the problem size.
3. Consider the Lagrangian relaxation of constraints (6b) without the aggregate capacity constraint (6d) added. Apply subgradient optimisation to compute approximate optimal multipliers $\overline{\eta}$ and let $(\overline{x}, \overline{y})$ denote the solution of the corresponding Lagrangian subproblem.

4. Try to find a partitioning $\{J_q : q \in Q\}$ of the depot set J such that $\sum_{j \in J_q} \overline{x}_{ij} > 1$ for at least one $i \in I$ and $q \in Q$. For this purpose apply the following steps:
   (a) Set $O = \{j \in J : \overline{y}_j = 1\}$ and $q = 0$.
   (b) Choose $i \in I$ such that $\sum_{j \in O} \overline{x}_{ij} > 1$. If there is no such $i \in I$, go to step (d).
   (c) Set $q := q + 1$, $J_q = \{j \in O : \overline{x}_{ij} > 0\}$, and $O := O \setminus J_q$. Go to step (b).
   (d) Assign each $j \in J$ with $\overline{y}_j = 0$ to the set J_q which minimises $\min_{l \in J_q} \sum_{i \in I} |c_{ij} - c_{il}|$.
5. Solve the primal linear master problem (20) by means of column generation. After each call to the Lagrangian subproblem (pricing subproblem) (21) solve the transportation problem if the set of open depots has enough capacity to meet total demand. Reapply Lagrangian probing if an improved feasible solution to the CFLP has been found.
6. Let μ^* denote the computed optimal solution of the linear program (20) and let (x^*, y^*), where $(y_j^*, x_{ij}^*) = \sum_{h \in H_q} \mu_{qh}^* (y_j^h, x_{ij}^h)$ for $j \in J_q$, denote the corresponding (fractional) solution in terms of the orginal variables. Round a fractional solution y^* and solve the resulting transportation problem in order to possibly obtain an improved feasible solution to the CFLP. Afterwards apply an interchange procedure to the best feasible solution obtained so far; however, do not allow the procedure to open (close) depots j if y_j^* is small (large).

A branch-and-price procedure has not been implemented. Possible branching rules are, however, to branch on a single variable y_j if y_j^* is fractional, or to impose the branching constraints $\sum_{j \in S} y_j = 0$ vs. $\sum_{j \in S} y_j \geq 1$ if $\sum_{j \in S} y_j^*$ is fractional. In order to perform the column generation, Klose and Drexl (2001) use the analytic center cutting plane method (Goffin et al. (1992)). The above procedure has been tested on 75 test problems ranging in size from 100 potential depot sites and 100 customers to 200 potential depot sites and 500 customers. The test problems differ in the ratio $r = \sum_j s_j / \sum_i d_i$ of total capacity to total demand. For each problem size and ratio $r \in \{3, 5, 10\}$ five problem instances have been generated according to a proposal of Cornuejols et al. (1991). Furthermore, the bound obtained by means of the partitioning approach has been compared with Lagrangian bounds based on relaxing the demand constraints and the capacity constraints in model (6), respectively. These last two bounds were also compute by means of stabilised column generation procedures. Since the partitioning approach usually only make sense for relatively large problem instance, we reproduce here the results obtained for the largest test problems with 200 depot sites and 500 customers (Table 1). The results shown are averages over the five instances of each problem type. In Table 1, $LB\,\%$ is the percentage deviation of the lower bound from optimality; $UB\,\%$ is the percentage deviation of the computed feasible solution from an optimal one; It_{LR} and It_M denote the number of Lagrangian

Table1. Computational results

r	$LB\,\%$	$UB\,\%$	It_{LR}	It_M	T_{LR}	T_H	T_M	T_{Tot}
Partitioning approach								
3	0.09	0.02	100	100	2514.2	9.6	154.8	2680.3
5	0.26	0.34	116	115	5657.0	7.1	217.9	5884.1
10	0.30	0.29	157	157	59076.4	3.5	1647.1	60727.9
"Conventional" relaxation of demand constraints								
3	0.16	0.02	300	24	8.9	17.0	9.9	36.7
5	0.40	0.37	631	68	25.7	21.2	39.6	87.6
10	0.47	1.09	901	91	42.1	35.5	172.3	251.7
Relaxation of capacity constraints								
3	0.15	0.00	290	290	1273.2	34.0	66.3	1373.5
5	0.33	0.04	267	267	11487.0	22.6	51.1	11560.9
10	0.29	0.02	190	190	9853.1	10.0	30.6	9893.9

subproblems and restricted linear master problems solved, respectively; T_{LR}, T_H, and T_M are the computation times spent on solving Lagrangian subproblems, computing heuristic solutions, and solving master problems; T_{Tot} is the total computation in seconds of CPU time on a Sun Ultra (300 MHz).

As can be seen from Table 1, the partitioning method produces strong lower bounds on the optimal solution value. For a number of large test problems this bound even improves the very strong bound based on relaxing the capacity constraints (6c). The computational effort required to solve the relaxation based on partitioning the depot set is, however, quite large; also the observed variation in the times spent on computing this bound was substantial. This is due to the complexity of the subproblem which itself decomposes into (smaller) CFLPs. A counterintuitive result is that the best lower bounds computed by means of the partitioning approach have been obtained for the test problems with smallest capacity tightness index r, although this relaxation method does not make use of the aggregate capacity constraint (6d). This indicates that the heuristic used for decomposing the depot set does not work well for this type of test problems and should be improved for problems with relatively loose capacity constraints. Nevertheless, due to the quality of the lower bounds, the method has some potential to solve large problems to optimality or at least very near to optimality. This is also shown by the results of an experiment with three single very large problem instances with 1000 customers and 500 potential depot sites (see Table 2). Since no optimal solution is known for these large instances, Table 2 only shows the percentage deviation GAP % between the upper and lower bound computed by means of the partitioning method.

Table2. Results for 3 single large-scale test problems

r	Gap%	It_{LR}	It_M	T_{LR}	T_H	T_M	T_{Tot}
3	0.06	129	129	21769.8	168.7	768.7	22735.1
5	0.10	167	166	43286.1	66.4	1392.1	44754.3
10	0.40	234	233	59007.7	49.4	2927.3	61997.1

5 Conclusions

In this paper, we discussed the application of a partitioning method to a number of optimisation problems of the assignment type, and compared this approach to the standard way of applying column generation to assignment problems. It has been shown, that the conventional way of transforming assignment type problems into problems of the set-partitioning type and solving the linear relaxation of this reformulation is equivalent to a Lagrangian relaxation of the semi-assignment constraints. This relaxation can be improved by imposing the constraints that no "activity" may be assigned more than once to a given subset of agents. Applying Lagrangian relaxation of the semi-assignment constraints while adding these implied constraints leads to a Lagrangian (pricing) subproblem which decomposes into smaller optimisation problems of the same type as the original optimisation problem. A necessary condition for obtaining this way an improved lower bound is, that the added implied constraints are "Lagrangian cuts" (see proposition 1). Computational results obtained with this approach for the CFLP indicate that the method is capable to solve large problem instances to optimality or very near to optimality. The partitioning principle is generally applicable to assignment problems which are decomposable if the semi-assignment constraints are relaxed. However, the method has in any case to be fine-tuned to the specific problem on hand. This raises a number of research questions concerning algorithmic design. Topics which have to be addressed are the design of heuristics for suitably decomposing the "agent" set J, the implementation of Lagrangian heuristics as well as primal heuristics using the information of a fractional solution to the primal master problem, the design of effective algorithms for solving the subproblems (which are of the same type as the original problem), the implementation of Lagrangian probing techniques for reducing the problem size, and finally the design of fine-tuned methods for stabilising the column generation.

Acknowledgement

This research was supported in part by the Swiss National Science Foundation (grant 12–63997).

References

Aardal, K. (1998): Capacitated facility location: Separation algorithm and computational experience. *Mathematical Programming*, 81:149–175.

Aardal, K. / Pochet, Y. / Wolsey, L. A. (1995): Capacitated facility location: Valid inequalities and facets. *Mathematics of Operations Research*, 20:552–582.

Aikens, C. H. (1985): Facility location models for distribution planning. *European Journal of Operational Research*, 22:263–279.

Arora, S. / Lund, C. (1997): Hardness of approximations. In: Hochbaum, D. S. (ed.), *Approximation Algorithms for NP-Hard Problems*, pp. 399–346. PWS Publishing Company, Boston.

Balachandran, V. (1976): An integer generalized transportation model for optimal job assignment in computer networks. *Operations Research*, 24:742–749.

Beasley, J. E. (1988): An algorithm for solving large capacitated warehouse location problems. *European Journal of Operational Research*, 33:314–325.

Bienstock, D. / Günlük, O. (1996): Capacitated network design – polyhedral structure and computation. *INFORMS Journal on Computing*, 8:243–259.

Boffey, T. B. (1989): Location problems arising in computer networks. *The Journal of the Operational Research Society*, 40:347–354.

Bruns, A. D. / Klose, A. / Stähly, P. (2000): Restructuring of Swiss parcel delivery services. *OR-Spektrum*, 22:285–302.

Cabot, A. V. / Erenguc, S. S. (1984): Some branch-and-bound procedures for fixed-cost transportation problems. *Naval Research Logistics*, 31:145–154.

Cabot, A. V. / Erenguc, S. S. (1986): Improved penalties for fixed cost linear programs using Lagrangean relaxation. *Management Science*, 32:856–869.

Campell, J. F. / Langevin, A. (1995): The snow disposal assignment problem. *The Journal of the Operational Research Society*, 46:919–929.

Caprara, A. / Salazar, J. J. G. (1999): Separating lifted odd-hole inequalities to solve the index selection problem. *Discrete Applied Mathematics*, 92:111–134.

Cattrysse, D. / Degraeve, Z. / Tistaert, J. (1998): Solving the generalised assignment problem using polyhedral results. *European Journal of Operational Research*, 108:618–628.

Cattrysse, D. / Salomon, M. / Van Wassenhove, L. N. (1994): A set-partitioning heuristic for the generalized assignment problem. *European Journal of Operational Research*, 72:167–174.

Cattrysse, D. / Van Wassenhove, L. N. (1992): A survey of algorithms for the generalized assignment problem. *European Journal of Operational Research*, 60:260–272.

Chardaire, P. (1999): Hierarchical two level location problems. In: Sansò, B. / Soriano, P. (eds.), *Telecommunications Network Planning*, chap. 3, pp. 33–54. Kluwer Academic Publishers Group, London, Dordrecht, Boston.

Christofides, N. / Beasley, J. E. (1983): Extensions to a Lagrangean relaxation approach for the capacitated warehouse location problem. *European Journal of Operational Research*, 12:19–28.

Cornuejols, G. / Sridharan, R. / Thizy, J.-M. (1991): A comparison of heuristics and relaxations for the capacitated plant location problem. *European Journal of Operational Research*, 50:280–297.

Crescenzi, P. / Kann, V. (1998): A compendium of NP optimization problems. In: Ausiello, G. / Crescenzi, P. / Gambosi, G. / Kann, V. / Spaccamela, A. M. C. / Protosi, M. (eds.), *Approximate Solution of NP-hard Optimization Problems*. Springer-Verlag, `http://www.nada.kth.se/~viggo/problemlist/compendium.html`.

Current, J. / Weber, C. (1994): Application of facility location modeling constructs to vendor selection problems. *European Journal of Operational Research*, 76:387–392.

Dantzig, G. B. / Wolfe, P. (1960): Decomposition principle for linear programs. *Operations Research*, 8:101–111.

Daskin, M. S. (1995): *Network and Discrete Location: Models, Algorithms, and Applications*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore.

Domschke, W. / Krispin, G. (1997): Location and layout planning: A survey. *OR-Spektrum*, 19:181–194.

Engeler, K. / Klose, A. / Stähly, P. (1999): A depot location-allocation problem of a food producer with an outsourcing option. In: Speranza, M. G./ Stähly, P. (eds.), *New Trends in Distribution Logistics*, vol 480 of *Lecture Notes in Economics and Mathematical Systems*, chap. 1, pp. 95–109. Springer-Verlag, Berlin, Heidelberg, New York.

Fisher, M. L. / Jaikumar, R. (1981): A generalized assignment heuristic for vehicle routing. *Networks*, 11:109–124.

Fisher, M. L. / Jaikumar, R. / van Wassenhove, L. N. (1986): A multiplier adjustment method for the generalized assignment problem. *Management Science*, pp. 1095–1103.

Fleischmann, B. (1990): The vehicle routing problem with multiple use of the vehicles. Working paper, Fachbereich Wirtschafts- und Organisationswissenschaften, Universität der Bundeswehr Hamburg, Hamburg.

Gelders, L. F. / Pintelon, L. M. / van Wassenhove, L. N. (1987): A location-allocation problem in a large Belgian brewery. *European Journal of Operational Research*, 28:196–206.

Geoffrion, A. M. / Graves, G. W. (1974): Multicommodity distribution system design by Benders decomposition. *Management Science*, 20:822–844.

Goffin, J.-L. / Haurie, A. / Vial, J.-P. (1992): Decomposition and nondifferentiable optimization with the projective algorithm. *Management Science*, 38:284–302.

Göthe-Lundgren, M. / Larsson, T. (1994): A set covering reformulation of the pure fixed charge transportation problem. *Discrete Applied Mathematics*, 48:245–259.

Guignard, M. / Kim, S. / Wang, X. (1997): Lagrangean probing in branch-and-bound. Technical Report 97-03-20, Operations and Information Management Department, The Wharton School, University of Pennsylvania.

Guignard, M. / Rosenwein, M. B. (1989): An improved dual based algorithm for the generalized assignment problem. *Operations Research*, 37:658–663.

Guignard, M. / Zhu, S. (1994): A two-phase dual algorithm for solving Lagrangean duals in mixed integer programming. Report 94-10-03, Operations and Information Management Department, University of Pennsylvania, The Wharton School.

Herer, Y. T. / Rosenblatt, M. J. / Hefter, I. (1996): Fast algorithms for single-sink fixed charge transportation problems with applications to manufacturing and transportation. *Transportation Science*, 30:276–290.

Hirsch, W. M. / Dantzig, G. B. (1968): The fixed charge transportation problem. *Naval Research Logistics*, 15:413–425.

Hochbaum, D. S. (ed.) (1997): *Approximation Algorithms for NP-Hard Problems*. PWS Publishing Company, Boston.

Hultberg, T. H. / Cardoso, D. M. (1997): The teacher assignment problem: A special case of the fixed charge transportation problem. *European Journal of Operational Research*, 101:463–473.

Jacobsen, S. K. (1983): Heuristics for the capacitated plant location model. *European Journal of Operational Research*, 12:253–261.

Karabakal, N. / Bean, J. / Lohmann, J. R. (1992): A steepest descent multiplier adjustment method for the generalized assignment method. Technical report 92-11, Department of Industrial and Operations Engineering, University of Michigan.

Kelley, J. E. (1960): The cutting-plane method for solving convex programs. *Journal of the SIAM*, 8:703–712.

Kennington, J. L. / Unger, E. (1976): A new branch-and-bound algorithm for the fixed charge transportation problems. *Management Science*, 22:1116–1126.

Khumawala, B. M. (1974): An efficient heuristic procedure for the capacitated warehouse location problem. *Naval Research Logistics Quarterly*, 21:609–623.

Klose, A. / Drexl, A. (2001): Lower bounds for the capacitated facility location problem based on column generation. Working paper, University of St. Gallen, Switzerland.

Kontoravdis, G. / Bard, J. F. (1995): A GRASP for the vehicle routing problem with time windows. *ORSA Journal on Computing*, 7:10–23.

Koopmans, T. C. / Beckmann, M. J. (1957): Assignment problems and the location of economic activities. *Econometrica*, 25:53–76.

Korupolu, M. R. / Plaxton, C. G. / Rajaraman, R. (1998): Analysis of a local search heuristic for facility location problems. Technical Report 98-30, DIMACS, Rutgers University.

Kuhn, H. (1995): A heuristic algorithm for the loading problem in flexible manufacturing systems. *International Journal of Flexible Manufacturing Systems*, 7:229–254.

Lee, D.-H. / Kim, Y.-D. (1998): A multi-period order selection problem in flexible manufacturing systems. *The Journal of the Operational Research Society*, 49:278–286.

Lemaréchal, C. (1989): Nondifferentiable optimization. In: Nemhauser, G. L. / Rinnooy Kan, A. H. G. / Todd, M. J. (eds.), *Optimization*, vol 1 of *Handbooks in Operations Research and Management Science*, pp. 529–572. North-Holland, Amsterdam.

Martello, S. / Toth, P. (eds.) (1990a): *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore.

Martello, S. / Toth, P. (1990b): Lower bounds and reduction procedures for the bin packing problem. *Discrete Applied Mathematics*, 28:59–70.

Mirchandani, P. B. / Francis, R. L. (eds.) (1990): *Discrete Location Theory*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore.

Mirzaian, A. (1985): Lagrangian relaxation for the star-star concentrator location problem: Approximation algorithm and bounds. *Networks*, 15:1–20.

Moore, E. W. / Warmke, J. M. / Gorban, L. R. (1991): The indispensible role of management science in centralizing freight operations at Reynolds metals company. *Interfaces*, 21:107–129.

Nauss, R. M. (1978): An improved algorithm for the capacitated facility location problem. *The Journal of the Operational Research Society*, 29:1195–1201.

Nemhauser, G. L. / Wolsey, L. A. (1988): *Integer and Combinatorial Optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore.

Osman, I. H. (1995): Heuristics for the generalized assignment problem: Simulated annealing and tabu search approaches. *OR-Spektrum*, 17:211–225.

Owen, S. H. / Daskin, M. S. (1998): Strategic facility location: A review. *European Journal of Operational Research*, 111:423–447.

Palekar, U. S. / Karwan, M. K. / Zionts, S. (1990): A branch-and-bound method for the fixed charge transportation problem. *Management Science*, 36:1092–1105.

Pirkul, H. (1986): An integer programming model for the allocation of databases in a distributed computer system. *European Journal of Operational Research*, 26:842–861.

Pochet, Y. / Wolsey, L. A. (1988): Lot-size models with backlogging: Strong reformulations and cutting planes. *Mathematical Programming*, 40:317–335.

Revelle, C. S. / Laporte, G. (1996): The plant location problem: New models and research prospects. *Operations Research*, 44:864–874.

Romero Morales, M. D. (2000): *Optimization Problems in Supply Chain Management*. Ph.D. thesis, Erasmus University, Rotterdam.

Ryu, C. / Guignard, M. (1992): An efficient algorithm for the capacitated plant location problem. Working Paper 92-11-02, Decision Sciences Department, University of Pennsylvania, The Wharton School.

Savelsbergh, M. W. (1997): A branch-and-price algorithm for the generalized assignment problem. *Operations Research*, 45:831–841.

Shmoys, D. B. / Tardos, E. / Aardal, K. (1997): Approximation algorithms for facility location problems. In: *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, pp. 265–274.

Sun, M. / Aronson, J. E. / McKeown, P. G. / Drinka, D. (1998): A tabu search heuristic procedure for the fixed charge transportation problem. *European Journal of Operational Research*, 106:441–456.

Taillard, É. D. / Laporte, G. / Gendreau, M. (1996): Vehicle routeing with multiple use of vehicles. *The Journal of the Operational Research Society*, 47:1065–1070.

Tüshaus, U. / Wittmann, S. (1998): Strategic logistic planning by means of simple plant location: A case study. In: Fleischmann, B. / van Nunen, J. A. E. E. / Speranza, M. G. / Stähly, P. (eds.), *Advances in Distribution Logistics*, vol 460 of *Lecture Notes in Economics and Mathematical Systems*, chap. 2, pp. 241–263. Springer-Verlag, Berlin, Heidelberg, New York.

Wentges, P. (1996): Accelerating Benders' decomposition for the capacitated facility location problem. *Mathematical Methods of Operations Research*, 44:267–290.

Wright, D. / Haehling von Lanzenauer, C. (1989): Solving the fixed charge problem with Lagrangian relaxation and cost allocation heuristics. *European Journal of Operational Research*, 42:305–312.

Chapter 5

Vehicle Routing and Transportation

The Vehicle Routing Problem with Time Windows and Simultaneous Pick-up and Delivery

Enrico Angelelli[1] and Renata Mansini[2]

[1] Department of Quantitative Methods , University of Brescia, Contrada S. Chiara 48/b, I-25122 Brescia, Italy, angele@eco.unibs.it
[2] Department of Electronics for Automation, University of Brescia, via Branze 38, I-25123 Brescia, Italy, rmansini@ing.unibs.it

Abstract. In this paper we consider the problem of a single depot distribution/collection system servicing a set of customers by means of a homogeneous fleet of vehicles. Each customer requires the simultaneous delivery and pick-up of products to be carried out by the same vehicle within a given time window. Products to be delivered are loaded at the depot and picked-up products are transported back to the depot. The objective is to minimize the overall distance traveled by the vehicles while servicing all the customers. To the best of our knowledge no exact algorithms have been introduced for this problem. We implement a Branch and Price approach based on a set covering formulation for the master problem. A relaxation of the elementary shortest path problem with time windows and capacity constraints is used as pricing problem. Branch and Bound is applied to obtain integer solutions. Known benchmark instances for the VRP with time windows have been properly modified to be used for the experimental analysis.

Keywords: Vehicle routing; Pick-up and Delivery; Branch-and-Price.

1 Introduction

In the last years the interest in reverse logistics received an increasing attention from companies involved in goods distribution and subsequent collection of used products for possible recycling. In a broadest sense reverse logistics is involved with all the activities concerning materials and products reuse. The high cost of refuse disposal along with the presence of environmental laws have forced firms to take care of their used products, while the challenge of finding an efficient/effective way to manage a reverse distribution network has become a critical issue for many distribution companies. In this paper we are interested in the integration of bi-directional flows in a single depot network transportation: the forward flow of products starting at a warehouse (delivery phase) with the backward flow of used products/waste ending to the warehouse (pick-up phase).

Past research about single depot routing problems with pick-up and delivery service has been mainly focused on two different types of problems:

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

- Pick-up and Delivery Problems (PDP). Products are picked up at some location and delivered to another location. Possible time windows can be added for each origin and destination. Two objective functions are typically considered: the number of vehicles needed to satisfy the requests and the total distance (or travel time). The Dial-a-ride problem, where persons have to be picked-up and transported to some destination, is a particular case of PDP with time windows (PDPTW). In this case the objective function is usually a measure of the inconvenience created by pick-ups or deliveries performed outside the desired time intervals.
- VRP with Backhauls (VRPB). Each customer has either a pick-up or a delivery demand to be satisfied. Products to be delivered are loaded at the depot while picked up products are transported back to the depot. A set of vehicle routes has to be designed so that all customers are serviced exactly once and no "pick-up customer" is visited before any other "delivery customer" on the same route. The objective is the minimization of the number of vehicles or the total length of the routes. A typical generalization of the problem takes time windows into account.

As far as exact algorithms are concerned, the first class of problems has been mainly studied with the addition of time window constraints. Sol and Savelsbergh (1995) have proposed an exact algorithm (Branch-and-Price) to solve the PDP with time windows (PDPTW) and tested it on instances with at most 50 customers. The branching strategies implemented by these authors make their algorithm more efficient with respect to the first exact algorithm proposed for the same problem by Dumas et al. (1991). Exact solution methods for the second class of problems can be find in Gelinas et al. (1995) and more recently in Mingozzi et al. (1999). A general survey on exact and heuristic solution methods for time constrained routing and scheduling problems can be found in Desrosiers et al. (1995).

In this paper we consider the problem dealing with a single depot distribution/collection system servicing a set of customers by means of a homogeneous fleet of vehicles. Each customer requires two types of service, a pick-up and a delivery. The critical feature of the problem is that both activities have to be carried out simultaneously by the same vehicle (each customer is visited exactly once). Products to be delivered are loaded at the depot and products picked-up are transported back to the depot. Moreover, each customer has to be visited within a given time window. The objective is to find the set of routes servicing all the customers at the minimum cost. This problem is a generalization of the Vehicle Routing Problem with Time Windows (VRPTW) since each customer requires a double service, by simultaneously delivering/picking up a given quantity of products. In the literature the most successful exact approaches for VRPTW are based on the solution of some relaxations of the shortest path problem. Desrochers et al. (1992) use a column generation scheme, Halse (1992) presents a decomposition based on variable splitting, while Kohl and Madsen (1997) develop an algorithm based

on Lagrangian relaxation. In all these cases the resulting subproblem is a shortest path problem with time windows and capacity constraints. More recently Kohl et al. (1999) propose for the VRPTW an effective strong valid inequality, the 2-path cut, which is a generalization of TSP subtour elimination constraints. The authors incorporate the cuts by adding extra rows to the master problem of a Dantzig-Wolfe decomposition approach. The authors are successful in solving to optimality previously unsolved problems.

From a practical point of view the VRPTW with simultaneous pick-up and delivery perfectly models all real situations (distribution of soft drinks, laundry service for restaurants and hotels) where the customers are typically visited only once, but for a double service. To the best of our knowledge no exact algorithms have been implemented for this problem, even if some suggestions can be found in Halse (1992), where the author proposes a mathematical formulation and a heuristic solution method. We implement a Branch and Price approach based on a set covering formulation of the master problem. A relaxation of the elementary shortest path problem with time windows and capacity constraints is used as pricing problem. Branch and Bound is applied to obtain integer solutions. Different branching strategies and some variants of a pricing algorithm are implemented in order to test their efficiency for this problem. The Solomon's test problems for the VRPTW has been used to generate modified benchmark instances for the experimental analysis.

The paper is organized as follows. In Section 2 the problem is described and the notation used is presented. In Section 3 a general mixed integer model for the problem as well as the set covering formulation for the column generation approach are presented. Section 4 is devoted to the pricing algorithm: different variants of a basic procedure are described as well as some re-optimization strategies. In Section 5 the Branch and Bound part of the procedure is analyzed. Finally, in Section 6 computational results are described and commented, while in Section 7 conclusions and future research are drawn.

2 Problem Description and Notation

Formally, the VRPTW with simultaneous pick-up and delivery is defined as follows. We consider a fleet of K homogeneous vehicles with equal capacity Q servicing a set N of customers, $N = \{1, 2, ..., n\}$, from/to a central depot (the unique depot is split into two identical nodes, indexed by 0 and $n + 1$, where the two indexes are used to emphasize the different role of the depot). Each customer i is characterized by his geographic location, his delivery and pick-up requests d_i and p_i, respectively and by the time window $[a_i, b_i]$ in which he must be serviced. Since each customer must be visited exactly once, it follows that $0 \leq d_i, p_i \leq Q, \forall i$. A vehicle is allowed to reach a customer before the opening of the time window and wait at no cost until the service is possible. In any case the vehicle is not permitted to arrive after the end of the

time window ("hard time windows"). The time window $[a_0, b_0] \equiv [a_{n+1}, b_{n+1}]$ associated to the central depot corresponds to the temporal horizon of the problem: it defines the availability of vehicles and/or drivers to the warehouse.

The problem belongs to the class of the *one to many and many to one* problem, as each vehicle must leave and return to the same depot, without any intermediate transhipment. A solution is a set of at most K routes starting and ending at the depot such that all the customers are visited exactly once and time and capacity constraints are satisfied. The objective is to minimize the total distance traveled by all the vehicles.

Let us denote by $G = (V, A)$ the complete graph induced by the customers, where $V = N \cup \{0, n+1\}$ is the set of nodes and A the set of arcs linking any pair of nodes. Let c_{ij} and t_{ij}, be the cost (the distance) and the duration (time required to cover the distance) associated to each arc $(i, j) \in A$, respectively. Without loss of generality, we include the service time at customer i into the duration of the arc (i, j). We assume that $a_i, b_i, d_i, p_i, Q, c_{ij}$ are non-negative integers, while t_{ij} are strictly positive integers. Finally, the triangular inequality holds both for times and costs, i.e. $c_{ik} + c_{kj} \geq c_{ij}$ and $t_{ik} + t_{kj} \geq t_{ij}$, $\forall i, k, j \in V$.

Let P be an elementary path in G, $P = \{0 = i_0, i_1, ..., i_p, i_{p+1} = n+1\}$, starting at vertex 0 and ending at vertex $n+1$. A feasible solution for our problem is represented by a set of disjoint elementary paths originating in 0 and ending in $n+1$. These paths altogether have to visit every customer exactly once, while satisfying the time windows and the capacity constraints. Notice that in our assumptions picked-up products cannot be used to satisfy delivery requirements. Thus, for every customer in a path, the pick-up demands already collected plus the quantities still to be delivered must not exceed the vehicle capacity.

Thus, a path P is feasible if, for each $s = 1, ..., p$, the following conditions hold:

$$a_{i_s} \leq T_{i_s} \leq b_{i_s} \tag{1}$$

$$\sum_{k=1}^{s} p_{i_k} + \sum_{k=s+1}^{p} d_{i_k} \leq Q, \tag{2}$$

where $T_{i_s} = \max\{a_{i_s}, T_{i_{s-1}} + t_{i_{s-1}, i_s}\}$ indicates the time at which service is started at node i_s.

3 Mathematical Formulation

As a generalization of the VRPTW (and thus also of the VRP), the VRPTW with simultaneous pick-up and delivery is NP-hard. Moreover, it can be easily shown that if the problem is a "net delivery", i.e. $p_i < d_i \ \forall i$, then the optimal

solution corresponds to the optimal solution of the VRP defined on the same graph G where each customer does require only the delivery service. In fact, if, for each customer, the delivery demand is higher than the corresponding pick-up quantity, the quantity collected will always be lower than the room made available by the corresponding delivery. Similar remarks hold for the case of a "net pick-up" ($p_i > d_i \ \forall i$).

In this section we present two different mathematical formulations for the VRPTW with simultaneous pick-up and delivery. The first one is a general mixed integer formulation, while the second one is the set covering formulation which will be used in the proposed column generation approach.

A general mixed integer programming model

Let us define the following four types of variables. For each arc (i, j) where $i \neq j$, $i \neq n+1$, $j \neq 0$ and each vehicle k, we define:

$X_{ij}^k = \begin{cases} 1 \text{ if the vehicle } k \text{ travels directly from } i \text{ to } j, \\ 0 \text{ otherwise;} \end{cases}$

D_i^k : the amount of the remaining deliveries carried by vehicle k when departing from customer i;

P_i^k : the amount of collected pick-up quantities carried by vehicle k when departing from customer i;

T_i^k : the starting time of the service of the vehicle k at customer i.

A mixed integer programming formulation for the VRPTW with simultaneous pick-up and delivery defined on the graph $G(V, A)$ is as follows:

$$Min \quad \sum_{k \in K} \sum_{(i,j) \in A} c_{ij} \, X_{ij}^k \tag{3}$$

$$\sum_{k \in K} \sum_{j \in V} X_{ij}^k = 1 \quad \forall i \in N \tag{4}$$

$$\sum_{i \in V} X_{ip}^k = \sum_{j \in V} X_{pj}^k \quad \forall p \in N, \ \forall k \in K \tag{5}$$

$$\sum_{j \in N} X_{0j}^k \leq 1 \quad \forall k \in K \tag{6}$$

$$\sum_{i \in N} X_{i,n+1}^k = \sum_{j \in N} X_{0j}^k \quad \forall k \in K \tag{7}$$

$$D_i^k + P_i^k \leq Q \quad \forall i \in V, \ \forall k \in K \tag{8}$$

$$D_{n+1}^k = 0 \quad \forall k \in K \tag{9}$$

$$D_0^k = \sum_{i \in N} \sum_{j \in N} X_{ij}^k \, d_i \qquad \forall k \in K \tag{10}$$

$$P_{n+1}^k = \sum_{i \in N} \sum_{j \in N} X_{ij}^k \, p_i \qquad \forall k \in K \tag{11}$$

$$P_0^k = 0 \qquad \forall k \in K \tag{12}$$

$$X_{ij}^k \, (P_i^k + p_j - P_j^k) = 0 \qquad \forall i, j \in V, \, \forall k \in K \tag{13}$$

$$X_{ij}^k \, (D_i^k - d_j - D_j^k) = 0 \qquad \forall i, j \in V, \forall k \in K \tag{14}$$

$$X_{ij}^k \, (T_i^k + t_{ij} - T_j^k) \leq 0 \qquad \forall i, j \in V, \, \forall k \in K \tag{15}$$

$$a_i \leq T_i \leq b_i \qquad \forall i \in V, \, \forall k \in K \tag{16}$$

$$D_i^k \geq 0 \qquad \forall i \in V, \, \forall k \in K \tag{17}$$

$$P_i^k \geq 0 \qquad \forall i \in V, \, \forall k \in K \tag{18}$$

$$X_{ij}^k \in \{0, 1\} \qquad \forall i, j \in V, \, \forall k \in K \tag{19}$$

The objective function (3) minimizes the total cost of the routes. Constraint set (4) states that each customer must be serviced by exactly one vehicle. Constraint set (5) guarantees that the vehicle entering and exiting from each node is the same. Notice that the constraint set:

$$\sum_{k \in K} \sum_{i \in V} X_{ij}^k = 1 \qquad j \in N \tag{20}$$

usually coupled with constraint set (4) in the assignment problem are redundant and thus excluded. The constraint sets (6) and (7) ensure that each vehicle is used at most once. In particular, the group of constraints (5), (6) and (7) are the so called flow constraints requiring that each vehicle leaves the depot (node 0) at most once, leaves a node p only if it has been visited and returns to the depot (node $n+1$) at most once. The constraint set (8) ensures that the onboard load of a vehicle k, when departing from a node i, has to be always lower than the vehicle capacity. The constraint sets (10) and (12) establish that each vehicle leaves the depot fully loaded with the products to be distributed, while the pick-up load is null. Conversely, the constraint sets (9) and (11) guarantee that when vehicles return back to the depot they

have distributed all their deliveries and are fully loaded with the picked-up quantities. The non linear sets of equations (13) and (14) establish that if arc (i,j) is visited by vehicle k, then the quantity to be delivered by the vehicle has to decrease by d_j while the quantity picked-up has to increase by p_j. The non linear constraint set (15) states that if vehicle k drives through arc (i,j), then the time at which the service will start at node j, will be greater than or equal to the time at which service started at node i plus the time to travel from i to j, i.e. $T_j^k \geq T_i^k + t_{ij}$. Notice that this type of constraints allows for waiting time at each node if the service time window is not open. The constraint set (16) sets the time window for each customer $i \in N$. Finally, (17) and (18) are nonnegative conditions, while (19) are binary constraints.

By keeping the binary property of the variables X_{ij}^k, constraints (13) can be linearized as follows:

$$P_j^k \geq P_i^k + p_j + Q\,(X_{ij}^k - 1) \qquad \forall i,j \in V,\ \forall k \in K \tag{21}$$

$$P_j^k \leq P_i^k + p_j - Q\,(X_{ij}^k - 1) \qquad \forall i,j \in V, \forall k \in K \tag{22}$$

Similarly for constraint set (14), while constraints (15) are linearized as:

$$T_j^k \geq T_i^k + t_{ij} - (1 - X_{ij}^k)\,T \qquad \forall i,j \in V, \forall k \in K \tag{23}$$

where T is a constant value arbitrarily large and Q is the vehicle capacity. Notice that such constraints generalize the subtour elimination constraints proposed by Miller, Tucker and Zemlin (1960) for the TSP.

A set covering formulation

The presented general mixed integer formulation has the main disadvantage of finding poor lower bounds. For this reason we have considered the following alternative set partitioning formulation:

Problem SP

$$\begin{aligned} \text{minimize} \quad & \textstyle\sum_{s \in S} c_s X_s \\ \text{subject to} \quad & \textstyle\sum_{s \in S} \delta_{is} X_s = 1, \quad \forall i \in N \\ & X_s \geq 0,\ \text{integer} \quad \forall s \in S \end{aligned}$$

where S is the set of all feasible paths in the graph G, δ_{is} is a constant with value 1 if route s visits customer i and 0 otherwise, while c_s is the cost associated to route $s \in S$, computed as the sum of the costs c_{ij} of the arcs making part of the route. The decision variable X_s, $s \in S$, is equal to 1 if

path $s \in S$ is selected and 0 otherwise. We do not need to set variable X_s as binary since it will never has value greater than 1.

Vehicle routing problems based on set partitioning formulations typically contain a huge number of variables and thus are solved by means of column generation techniques.

The original set partitioning problem is usually identified as Master Problem (MP). The main idea is to solve a restricted formulation of the MP – the Restricted Master Problem (RMP) – characterized by a smaller number of variables to be more efficiently handled. Given the optimal solution of a RMP, a Pricing Problem (PP) is solved to find out columns with negative reduced costs candidate to enter the basis. If new columns are identified, they are added to the RMP which is then re-optimized, otherwise the current solution is optimal also for the original problem.

In our case, this column generation approach can be seen as equivalent to Dantzig-Wolfe decomposition: the problem decomposes into a master problem which selects the paths such that customers are serviced exactly once (it is equivalent to constraint (4)), while a subproblem generates new paths using the optimal solution of the dual of the restricted master problem by solving an elementary shortest path problem with time and simultaneous pick-up/delivery conditions (constraints (5) – (23)).

If the linear relaxation solution of the MP is not integer a branching strategy can be applied to close the integrality gap. The combination of the column generation approach with a branching scheme defines the so called Branch and Price algorithm.

Although the discussion above has considered a set partitioning formulation, in our implementation the master problem has a set covering type formulation. The latter has the advantage of a numerically more stable linear programming relaxation:

$$\text{minimize} \sum_{s \in S} c_s X_s \tag{24}$$

$$\text{subject to} \sum_{s \in S} \gamma_{is} X_s \geq 1, \quad \forall i \in N \tag{25}$$

$$\sum_{s \in S} X_s - X_d = 0 \tag{26}$$

$$\sum_{s \in S} c_s X_s - X_c = 0 \tag{27}$$

$$X_s \geq 0, \text{ integer} \quad \forall s \in S \tag{28}$$

$$X_d \in [1, K], \text{ integer} \tag{29}$$

$$X_c \geq 0 \text{ integer} \tag{30}$$

In this formulation columns correspond to routes and rows correspond to the requirement that each customer is visited *at least once* (γ_{is} can be any integer and corresponds to the number of time route s visits customer i).

As it will be thoroughly explained in Section 4, the fact that coefficients γ_{is} can be any integer larger than 1 facilitates the solution of the pricing problem. Notice that, in this case, a node is visited more than once by the same vehicle. The present formulation is less constrained with respect to the 0-1 set covering one, but its linear relaxation can be solved with much less effort.

Triangular inequality on times and costs guarantees that the optimal solution to the set covering problem is also optimal to the set partitioning.

In our set covering formulation of the master problem we have added two additional integer variables X_d and X_c. The variable X_d represents the number of employed vehicles and we use it for branching. The variable X_c is clearly the value of the objective function and is guaranteed to be integer at the optimum if all c_{ij}, $\forall (i,j) \in A$ are integer. Thus, a cut on this variable can be added whenever a LP relaxation, throughout the branch and bound tree, finds a fractional value of the objective function.

4 The Pricing Problem

Given the linear relaxation solution of the RMP, the pricing problem seeks for some columns in the MP, if there are any, with negative reduced costs. If no columns with negative reduced costs exist, the current solution of the relaxed RMP is also optimal for the relaxed MP.

The pricing problem is formulated as follows.

Let $(\pi, \lambda_d, \lambda_c)$ be the optimal solution of the dual of the current RMP, where π is the dual vector associated with covering constraints (25) and λ_d and λ_c the dual variables associated to the constraints (26) and (27), respectively.

From linear programming the reduced cost $\bar{c}_s$ for a path s is given by:

$$\bar{c}_s = c_s - \sum_{i \in N} \pi_i \gamma_{is} - \lambda_d - \lambda_c c_s.$$

If any negative reduced costs exist, then we have identified columns to enter the basis and the RMP is reoptimized; otherwise we have proved that the current solution of the RMP is also optimal for the MP.

The cost of a path can be computed as the sum of the arcs it is made of. Thus, given the path visiting p nodes $(0 = i_0, i_1, \ldots, i_p, i_{p+1} = n+1)$, its cost will be equal to $c_s = \sum_{r=0}^{p} c_{i_r, i_{r+1}}$. Thus, $\bar{c}_s$ can be reformulated as:

$$\bar{c}_s = \sum_{r=0}^{p} c_{i_r, i_{r+1}} - \sum_{r=1}^{p} \pi_{i_r} - \lambda_d - \lambda_c \sum_{r=0}^{p} c_{i_r, i_{r+1}}$$

from which

$$\bar{c}_s = \sum_{r=0}^{p} [(1-\lambda_c)c_{i_r,i_{r+1}} - \pi_{i_r}], \text{ when we set } \pi_0 := \lambda_d.$$

Then, if we define the marginal cost $\bar{c}_{ij}$ of the arc (i,j) as:

$$\bar{c}_{ij} = (1-\lambda_c)c_{ij} - \pi_i,$$

finding a path with negative reduced cost is equivalent to determine a path with negative length in the graph $\overline{G} = (V, \overline{A})$ derived from the original one with modified costs $\bar{c}_{ij}$. On the other hand, proving the optimality of the current solution of the RMP requires the determination of the shortest path.

Although we are interested to find elementary paths, it has been shown (see Dror (1994)) that the elementary shortest path problem with resource constraints (ESPR) is Np-hard in the strong sense and no efficient algorithms are available to solve it.

The choice of a set covering formulation of the master problem which allows the selection of routes (columns) in which a customer may be visited more than once, permits to formulate the pricing problem as a Shortest Path Problem with resource constraints (SPPR) without the elementary condition.

The finiteness of feasible paths is guaranteed by time windows and capacity constraints and the SPPR remains NP-hard due to possible negative costs associated to the arcs; nevertheless, pseudo-polynomial algorithms based on dynamic programming exist for the problem (see Desrosiers et al. (1995)).

To solve our SPPR we have implemented a permanent labeling algorithm, by extending the procedure proposed by Desrochers and Soumis (1988b) for the SPP with time windows to the case with simultaneous pick-up and delivery demands.

Generally, a permanent labeling algorithm assigns tentative labels to nodes at each step, designating one (or more) label as permanent at each iteration. The algorithm selects a temporary label associated to a node i, makes it permanent and scans arcs in $A(i)$ (where $A(i)$ is the set of arcs emanating from node i) to generate new temporary labels of the adjacent nodes. The algorithm terminates when it has designated as permanent all the labels. The shortest path will be the one associated to the label of node $n+1$ with minimum cost.

To each node i in a path k is associated the following label:

$$(T_i^k, C_i^k, AL_i^k, ML_i^k),$$

where T_i^k is the time at which service is started at node i, C_i^k, is the cost of the path up to i, AL_i^k is the sum of the pick-up loads collected up to node i included and ML_i^k is the minimum capacity required to service all the nodes up to node i included. Obviously, for any path k the label $(T_0^k, C_0^k, AL_0^k, ML_0^k) \equiv (0,0,0,0)$. All the other labels are generated iteratively along the path as follows:

$$T_j^k := \max\{a_j, T_i^k + t_{ij}\}$$
$$C_j^k := C_i^k + \bar{c}_{ij}$$
$$AL_j^k := AL_i^k + p_j$$
$$ML_j^k := \max\{AL_j^k, ML_i^k + d_j\}.$$

The new label in node j, $j \in N$, is created only if feasible, i.e. if $T_j \leq b_j$ and $ML_j \leq Q$.

Definition 1 *Given two labels* $(T_i^k, C_i^k, AL_i^k, ML_i^k)$ *and* $(T_i^h, C_i^h, AL_i^h, ML_i^h)$ *associated to the same node* i, *the* **label** $(T_i^k, C_i^k, AL_i^k, ML_i^k)$ *is said to be* **dominant** *with respect to the label* $(T_i^h, C_i^h, AL_i^h, ML_i^h)$, *if and only if* $(T_i^h - T_i^k) \geq 0$ *and* $(C_i^h - C_i^k) \geq 0$ *and* $(AL_i^h - AL_i^k) \geq 0$ *and* $(ML_i^h - ML_i^k) \geq 0$. *Similarly, path* k *is said to dominate path* h *at node* i.

Definition 2 *A label* $(T_i^k, C_i^k, AL_i^k, ML_i^k)$, *is called* **efficient** *if it is not dominated by any other label* $(T_i^h, C_i^h, AL_i^h, ML_i^h)$. *Similarly, we call* **efficient** *the path* k *at node* i *if the corresponding label is efficient.*

The reported dominance relation is not a total ordering. This implies that more than one efficient paths, corresponding to non dominated labels, can be assigned to the same node.

The criteria used, at each iteration, to create new labels and to select temporary labels to be made permanent is critical to algorithm efficiency. The SPPR creates labels only if feasible and efficient. Moreover, the algorithm deletes temporary labels dominated by the newly created ones; finally, it selects and makes labels permanent according to a qualification condition such that a new label will never dominate a label which has been previously treated.

An algorithm for the SPPR may select and make permanent the label with smallest time. Since arc times are strictly positive, new generated labels cannot dominate the permanent ones.

In order to efficiently manage labels selection we have used buckets similarly to what has been done in Desrochers and Soumis (1988b). Buckets are time intervals which partition the time axis. The size of a bucket is set equal to the smallest arc duration, i.e.

$$h := \min\{t_{ij}\}, \quad \forall (i,j) \in A;$$

the p-th bucket will contain all the labels, possibly associated to different nodes, whose field T is included in the interval $[ph, (p+1)h[$. The number of activated buckets depends on depot time window size $[a_0, b_0]$, while the number of labels for each bucket is a function of the number of nodes and of the number of labels generated in each node.

In the label selection step we scan the buckets in time increasing order until we identify the first nonempty bucket. One by one we delete the labels into the selected bucket by making them permanent and scanning their adjacent nodes to generate new labels belonging to higher numbering buckets. A new label at node j is created only if the time and capacity constraints are satisfied: $T_j \leq b_j$ and $ML_j \leq Q$. The new generated labels will be compared with the other temporary labels to test possible dominance and to eliminate dominated labels.

In the following we briefly describe the basic version of the implemented SPPR Algorithm.

SPPR Algorithm

1 **begin**
//Determination of buckets size //
2 $dim\_bck := \min\{t_{ij}\} \forall (i,j) \in A$
//Determination of number of buckets //
3 $num\_bck := int(\frac{b_{n+1}}{dim\_bck})$
4 Create and initialize necessary buckets indexed from 0 to $num\_bck - 1$
// Setting of the initial label//
5 Set $T_0 := 0$, $C_0 := 0$, $AL_0 := 0$, $ML_0 := 0$ and assign it to bucket 0
6 **for p** $= 0$ **to** $num\_bck - 1$ scan the buckets with temporary labels
7 **while** bck_p is not empty
8 Cancel label (T_i, C_i, AL_i, ML_i) from the bucket
9 Save current label as permanent
10 **for** each successor j of node i
// Determination of the fields of the new label//
11 $T_j := \max\{T_j, a_j\}$
12 $AL_j := AL_i + p_j$
13 $ML_j := \max\{AL_j, ML_i + d_j\}$
//Test on time and capacity constraints feasibility //
14 **if** $T_j \leq b_j$ AND $ML_j \leq Q$
15 $C_j := C_i + \bar{c}_{ij}$
16 Create the label $L_j := (T_j, C_j, AL_j, ML_j)$
17 Determine the bucket into which the label has to be inserted
// Apply dominance test on the existing labels //
18 **if** label L_j is not dominated by permanent labels
19 Insert L_j into the appropriate bucket
20 Eliminate temporary labels dominated by L_j
21 **end** // if label is not dominated//
22 **end** //if $T_j \leq b_j$ AND $ML_j \leq Q$//
23 **end** //for each successor j//
24 **end** //while the bucket bck_p is not empty//
25 **next p**
26 **end** //end algorithm//

4.1 Pricing Strategies

As already stated, the pricing problem consists either in finding columns with negative reduced costs, if the are any, or in proving that no such columns exist. Notice that any column with negative reduced cost is a candidate to enter the basis of the current RMP for re-optimization. Thus, as soon as one candidate column is available, we do not need to find the shortest path and the algorithm can be stopped earlier.

For the generation of columns we have applied a 2-loop elimination strategy which was first introduced by Houck et al. (1980). Such strategy returns paths which do not contain cycles of the type $i - j - i$. Notice that, in this case, the basic version of the SPPR algorithm becomes more complex requiring labels with additional fields.

We have implemented three different pricing strategies. The basic strategy executes the SPPR algorithm giving as output all the paths with negative length. If the number of generated paths is larger than a fixed threshold the algorithm is stopped earlier.

The remaining two strategies are re-optimization procedures. The basic idea of these re-optimizations relies on the fact that to solve the VRPTW with simultaneous *pick-up* and *delivery* we need for disjoint paths in which each customer is visited by exactly one vehicle. A similar procedure was proposed by Desrochers and Soumis (1988a) for the shortest path problem with time windows.

There are different ways to get paths which do not intersect. The first re-optimization procedure works as follows. Once found a path with negative cost, its nodes are deleted from the graph and the procedure is iterated until either the graph is empty or no negative path exists. The stored disjoint paths represent the set of columns passed to the RMP.

The second re-optimization, first finds all the negative paths, then stores a subset of disjoint ones and finally eliminates their nodes from the graph. The procedure is iterated analogously to the first one.

5 The Branching Strategy

The column generation scheme usually terminates without finding an integer solution for the Set Covering problem. In this case, if the total distance traveled by the vehicles is equal to a fractional value c, we impose the following cut: $X_c \geq \lceil c \rceil$. The dual variable λ_c of the added constraint is correctly transferred to the pricing problem.

When, however the optimal solution of the relaxed problem is integer and it does not contain paths with cycle (this is always the case if the costs satisfy the triangle inequality) this is also the optimal solution for the VRPTW with simultaneous pick-up and delivery.

In all the other cases the integrality gap is closed by Branch and Bound. The use of effective branching strategies is a critical aspect of a Branch and

Price scheme: the new constraints, introduced with the branching strategies, must be compatible with the pricing problem, which in our case must remain a SPPR.

We have implemented two types of branching rules which do not modify the pricing problem. The first class takes into account the variable representing the number of vehicles as possible branching variable. In the optimal solution this variable must be integer. In the set covering formulation of the problem we have indicated this variable as X_d. If the optimal solution contains a fractional number v, of vehicles two branches will be created: in the first one we will set $X_d \leq \lfloor v \rfloor$ in the second one $X_d \geq \lceil v \rceil$. In both cases the dual variable λ_d associated to the new constraint is correctly added to the pricing problem, as shown in Section 3.

In order to always be able to restore problem feasibility we have added two dummy variables in the set covering formulation. A high cost is associated to such variables so that, after being used to restore feasibility, they leave the basis.

Notice that, in this class of strategies, we do not take into account the direct branching on variables X_s representing the route. In fact, while we can easily set a fractional variable X_s to 1 (all the arcs incident to the nodes belonging to the path are removed) it is difficult to set the same variable to 0.

The second class of branching strategies takes into account the arcs of the graph and applies all the times variables X_c and X_d are integer, but some variables X_s are still fractional. We first select a path which contains cycles. If such path exists we identify the first node visited more than once and we branch on the first arc incident on it. If no routes contain cycles, we look for a node shared by two routes either entering the node from two different predecessors or leaving the node directed to two distinct successors. For such a node we are able to select an arc belonging to only one of the two routes and branch on it.

Branching on an arc (i, j) means to set this arc respectively to 1 and to 0. In the first case, the arcs $(i, k) \in A$, $k \neq j$ and $(l, j) \in A$, $l \neq i$ are eliminated from the graph and all the routes using theses arcs are eliminated from the subproblem. When an arc is set to 0, we remove it from the graph and all the routes using the arc are eliminated from the subproblem.

6 The Computational Results

In this section we present and discuss the computational results. The main objective of these experiments is to test the efficiency of the different proposed strategies.

The Branch and Price algorithm has been implemented in Visual C++ using ABACUS, a specialized software framework for developing Branch and Price and Branch and Cut algorithms. The experiments have been conducted on a PC Pentium III with 800 MHz.

Since the VRPTW with simultaneous pick-up and delivery has never been solved before, no instances are available in the literature to be used for testing. For this reason we decided to test our Branch and Price algorithm on Solomon's instances for the VRPTW by modifying them in order to take into account the presence of simultaneous quantities of pick-up and delivery. In particular, the quantities given in the original Solomon's instances have been assumed to be the delivery demands d_i, $i = 1, ..., n$. Then, the pick-up demands p_i, $i = 1, .., n$ have been generated as follows:

$$p_i := \begin{cases} \lfloor (1-\alpha) d_i \rfloor \text{ if } i \text{ is even;} \\ \lfloor (1+\alpha) d_i \rfloor \text{ if } i \text{ is odd;} \end{cases}$$

where $0 \leq \alpha \leq 1$. Moreover, the vehicle capacity has been set equal to half of its original value.

The experiments have shown that the optimal solutions are not affected by the presence of the pick-up demands, unless we reduce the capacity of the vehicles. This is especially true when the number of customers taken into account is small.

The Branch and Price algorithm has been tested on the sets of Solomon problems in which customers are located in clusters (C problems) or randomly (R problems) and on problems based on a mixed structure of random and clustered customers (RC problems). For all these sets we have taken into account the first 20 customers. Euclidean distances among customers have been approximated to the first decimal digit. Then, we have added a scalar equal to 0.1 to all arcs costs c_{ij} to ensure that the triangular inequality on costs was satisfied (this value has been subsequently subtracted from the solutions obtained). The initial fractional values have been multiplied by 10 to get integer numbers.

Testing the algorithm has shown the importance of dominance test in drastically reducing the number of generated labels and thus the computational time.

The two proposed re-optimization strategies turned out to be unefficient, always requiring high computational time. A possible reason is due to the fact that, if the first columns generated by the pricing algorithm are of poor quality, then this will prevent the determination of more promising routes passing through the eliminated nodes. It could be interesting, in future, to relax this idea of re-optimization by allowing a partial overlapping of the routes returned by the pricing algorithm.

We have tested different versions of the algorithm. The SPPR has been solved with and without 2-loop elimination and with different thresholds on the number of columns returned by the pricing algorithm to the RMP.

In particular, we have reported the results found when 2-loop elimination is applied and the threshold is set to 100 (version 1) and 200 (version 2), respectively. See Table 1 and Table 2. In both cases the branch and bound uses a depth first strategy of backtracking.

Table 1. Computational results for $\alpha = 0.2$

			CPU (sec.)		No. Pricing		Columns		Pricing (sec.)		nSub		Tree level	
Problem	Cost	Vehicles	1	2	1	2	1	2	1	2	1	2	1	2
c101	2737	4	2.35	2.06	290	279	6665	5932	1.4	1.21	57	61	17	19
c102	2719	4	55.96	44.98	2406	1844	104825	104164	45.03	35.69	456	396	31	27
c103	2695	4	129.77	123.11	2087	1937	85844	113227	121.11	113.72	418	411	29	29
c104	2695	4	390.40	337.67	4468	4332	214493	301725	370.08	314.39	787	829	35	37
c105	2737	4	2.28	2.46	266	263	6139	7487	1.44	1.5	53	49	21	18
c106	2737	4	1.98	0.97	253	94	5394	2583	1.24	0.67	51	20	20	11
c107	2737	4	70.38	4.11	9098	438	195454	11935	42.61	2.74	1680	87	40	24
c108	2730	4	12.76	10.98	902	762	27667	28455	9.42	8.11	177	154	24	28
c109	2713	4	57.28	85.69	1709	4556	72871	203182	49.95	66.33	320	1054	35	42
r101	5103	7	0.03	0.03	4	4	113	113	0.01	0.02	1	1	1	1
r102	4340	6	0.14	0.12	21	11	455	552	0.1	0.05	3	3	2	2
r103	3705	5	3.16	3.59	125	211	5608	10852	2.72	2.92	23	38	14	11
r104	3454	3	17.08	43.22	228	642	12998	55146	16.09	39.23	31	82	12	21
r105	4317	5	0.05	0.06	7	9	215	270	0.02	0.04	1	1	1	1
r106	3838	4	0.74	0.47	48	37	2095	1755	0.61	0.36	8	6	5	4
r107	3491	4	0.42	0.26	23	12	1884	1533	0.37	0.16	1	1	1	1
r108	3199	3	0.37	0.51	15	17	1229	2069	0.34	0.43	1	1	1	1
r109	3724	4	0.08	0.09	8	7	359	473	0.05	0.06	1	1	1	1
r110	3608	4	1.59	1.35	34	35	1821	2411	1.42	1.18	4	4	3	3
r111	3530	4	0.21	0.24	14	13	939	1130	0.16	0.19	1	1	1	1
r112	3234	3	10.64	12.29	134	126	7859	11710	10.08	11.43	13	16	7	9
rc101	4416	5	0.19	0.17	24	19	599	634	0.1	0.09	4	4	3	3
rc102	4310	5	0.64	0.66	26	22	1132	1125	0.53	0.59	3	3	2	2
rc103	4285	5	2.45	1.68	69	55	3320	2894	2.21	1.47	17	13	8	7
rc104	4227	5	2.87	1.29	49	25	2870	1723	2.67	1.17	5	3	3	2
rc105	4325	5	0.73	0.6	35	26	1631	1349	0.61	0.49	4	4	3	3
rc106	4274	5	0.63	0.55	29	22	1469	1488	0.49	0.44	4	4	3	3
rc107	4195	5	1.79	2.17	50	53	2308	3845	1.62	1.91	8	8	5	5
rc108	4170	5	3.76	4.2	65	54	3298	3587	3.53	3.96	11	11	7	7
average			26.58	23.64	775.41	548.45	26605.31	30460.31	23.66	21.05	142.86	112.62	11.55	11.14

We have also tested the efficiency of a possible reuse of the generated columns. However, in a first group of computational results we have noticed that the number of columns actually reused was very small and thus we have decided to abandon the strategy.

Tables 1 and 2 report the details on the computational results for a value of the parameter α equal to 0.2 and 0.8, respectively.

In each table the first column refers to the problem solved, the second and third give the total cost of the routes and the number of vehicles used, respectively. The columns from fourth to ninth are divided into two parts according to the two implemented versions identified by numbers 1 and 2. Column CPU gives the CPU times expressed in seconds and include I/O operations and all the overheads. The columns named "No. Pricing" and "Columns" give the number of times pricing algorithm is called and the number of columns generated, respectively. The last three columns report the CPU time required by the pricing algorithm (Pricing), the number of nodes of the branch and bound tree which have been explored (nSub) and the length - as number of levels - of the generated tree (Tree level).

By comparing the results reported in the two tables it seems that, on average, the following observations can be drawn:

1. the number of generated columns always increases when passing from version 1 to version 2;
2. the number of generated subproblems in the branch and bound tree decreases in moving from $\alpha = 0.2$ to $\alpha = 0.8$ and from version 1 to version 2;

Table 2. Computational results for $\alpha = 0.8$

			CPU (sec.)		No. Pricing		Columns		Pricing (sec.)		nSub		Tree level	
Problem	Cost	Vehicles	1	2	1	2	1	2	1	2	1	2	1	2
c101	2862	5	3.85	1.91	405	194	7826	4743	2.73	1.34	61	30	23	13
c102	2800	4	3.24	2.87	70	56	4403	5724	2.99	2.61	9	7	6	5
c103	2745	4	11.98	8.48	243	144	12057	12149	10.84	7.8	40	22	13	9
c104	2743	4	87.45	188.49	778	663	39886	50434	83.98	184.6	125	120	21	21
c105	2862	5	5.24	1.06	556	79	11809	2808	3.51	0.75	100	11	26	6
c106	2862	5	3.97	0.62	422	55	8559	1788	2.86	0.4	73	9	20	6
c107	2798	4	0.26	0.26	21	19	841	1147	0.17	0.14	1	1	1	1
c108	2798	4	2.27	3.72	115	245	5471	10033	1.76	2.72	20	49	13	14
c109	2798	4	13.96	13.07	479	456	18969	24787	12.14	11.21	87	90	26	25
r101	5103	7	0.04	0.04	4	4	119	119	0	0.01	1	1	1	1
r102	4340	6	0.14	0.15	11	19	439	595	0.05	0.08	3	3	2	2
r103	3762	5	0.5	0.87	25	35	1154	2349	0.4	0.67	4	6	3	4
r104	3518	4	53.5	39.27	294	177	17298	15894	51.87	38.05	43	29	15	13
r105	4317	5	0.1	0.06	8	8	206	271	0.03	0.03	1	1	1	1
r106	3883	4	0.71	0.71	46	40	1814	2549	0.57	0.56	7	7	5	5
r107	3555	4	0.28	0.26	16	13	985	1132	0.2	0.21	1	1	1	1
r108	3474	3	95.16	66.77	398	324	20470	26751	93.18	64.99	68	55	20	15
r109	3724	4	0.11	0.08	9	6	466	462	0.07	0.05	1	1	1	1
r110	3677	4	2.69	2.5	37	30	1934	2030	2.45	2.33	4	4	3	3
r111	3533	4	0.2	0.19	10	9	742	1017	0.16	0.15	1	1	1	1
r112	3479	4	16.67	15.29	314	280	14509	17329	15.3	14.08	60	58	17	16
rc101	4601	5	0.07	0.06	9	6	342	294	0.04	0.03	1	1	1	1
rc102	4317	5	0.16	0.14	9	9	568	652	0.08	0.09	1	1	1	1
rc103	4292	5	0.25	0.2	12	10	875	805	0.15	0.16	1	1	1	1
rc104	4234	5	0.22	0.24	10	9	637	976	0.17	0.21	1	1	1	1
rc105	4493	5	0.15	0.15	11	12	546	784	0.08	0.11	1	1	1	1
rc106	4427	5	0.18	0.15	10	11	441	829	0.08	0.1	1	1	1	1
rc107	4253	5	0.27	0.27	10	9	746	766	0.16	0.21	1	1	1	1
rc108	4180	5	1.03	0.93	16	15	982	1477	0.92	0.86	3	3	2	2
average			10.51	12.03	149.93	101.28	6037.72	6575.66	9.89	11.54	24.83	17.79	7.86	5.93

3. on average version 1 solves much more LPs, however the CPU time for the two versions has the same order of magnitude;
4. the instances with $\alpha = 0.2$ are much more difficult of those generated with $\alpha = 0.8$ probably due to a less constraining capacity (note that for $\alpha = 0$ the instances reduce to VRPTW).

7 Conclusions and Future Research

In this paper we present a Branch and Price approach to solve a generalization of the VRPTW taking into account simultaneous pick-up and delivery demands for each customer. We have implemented different branching and pricing strategies testing their efficiency. Computational study has been conducted on instances obtained by modifying Solomon benchmark problems for the VRPTW. In particular we have compared two different versions of the pricing algorithm fixing the maximum number of routes returned to 100 and 200. It is interesting to notice that for any value of the parameter α the CPU time required by the two versions is independent by the number of subproblems. This is mainly due to the fact that the fraction of time spent in solving the relaxation is on average negligible.

As future research, some interesting implementation issues are:

1. to find some heuristics which eliminate cycles in the columns generated by the pricing algorithm and possibly return better paths to the RMP;
2. about pricing strategies, to introduce a different re-optimization procedure which better exploits the solution found at each iteration and tries to reduce the computational burden of the successive iterations;

3. about branching strategies, to consider the introduction of branching strategies on time windows and possibly the extension of 2-path cuts proposed in Kohl et al. (1999) for VRPTW to our problem.

References

Desrochers M., Desrosiers J., Solomon M. (1992): A new optimization algorithm for the vehicle routing problem with time windows. *Operations Research* 40 (2), 342–354.

Desrochers M., Soumis F. (1988): A reoptimization algorithm for the Shortest Path Problem with Time Windows. *European Journal of Operational Research*, 35, 242–254.

Desrochers M., Soumis F. (1988): A generalized permament labeling algorithm for the Shortest Path Problem with Time Windows. *INFOR* 26 193–214.

Desrosiers, J., Dumas,Y., Solomon, M.M., Soumis, F. (1995): Time constrained routing and scheduling. M.O. Ball, T.L. Magnanti, C.L. Monma, G.L. Nemhauser (eds.). *Networks Routing*, Handbooks in Operations Research and Management Science 8, North-Holland, Amsterdam, 35 - 139.

Dror M. (1994): Note on the complexity of the shortest path models for column generation in VRPTW, *Operations Research* 42(5) 977–978.

Dumas Y., Desrosiers J. e Soumis F. (1991): The Pickup and Delivery Problem with Time Windows. *European Journal of Operational Research* 54, 7–22.

Gelinas S., Desrochers M., Desrosiers J., Solomon M.M. (1995): A new branching strategy for time constrained routing problems with application to backhauling. *Annals of Operations Research* 61, 91–109.

Halse K. (1992): *Modeling and solving complex vehicle routing problems.* PhD Thesis no.60, IMSOR, The Technical University of Denmark.

Houck D.J., Picar, J.C., Queyranne and Vemuganti R.R., (1980): The Travelling Salesman Problem as a Constrained Shortest Path Problem: Theory and Computational Experience, *Opsearch*, 17, 93–109.

Kohl N., Madsen O. (1997): An optimization algorithm for the Vehicle Routing Problem with Time Windows based on Lagrangian Relaxation, *Operations Research*, 45(3), 395–406.

Kohl N., Desrosiers J., Madsen O., Solomon M., Soumis F. (1999): 2-path Cuts for the Vehicle Routing Problem with Time Windows, *Transportation Science*, 33(1), 101–116.

Miller C., Tucker A. and Zemlin R. (1960): Integer programming formulation and travelling salesman problems. *Journal of ACM* 7, 326 – 329.

Mingozzi A., Baldacci R., Giorgi S. (1999): An exact method for the Vehicle Routing Problem with Backhauls, *Transportation Science* 33(3), 315–329.

Sol M., Savelsbergh M.W.P. (1995): A Branch and Price Algorithm for the Pickup and Delivery Problem with Time Windows. *Report of Computational Optimization Center, Georgia Institute of Technology, Atlanta, Georgia.*

The Application of a Vehicle Routing Model to a Waste Collection Problem: Two Case Studies

Enrico Angelelli and Maria G. Speranza

Department of Quantitative Methods , University of Brescia, Contrada S. Chiara 48/b, I-25122 Brescia, Italy, angele,speranza@eco.unibs.it

Abstract. In this paper we propose a unique model for the estimation of the operational costs of each of three waste collection systems. In the traditional system, widely used, the waste is typically collected in plastic bags and a three men crew is needed on each vehicle. Two other systems, which require one man crew per vehicle and collect street containers, are considered. The side-loader system with fixed body automatically empties the street containers into the vehicle body and empties the body at the disposal site. The side-loader system with demountable body allows the separation of the waste collection phase from the transportation to the disposal site, since the vehicle body can be demounted. We also present two case studies and show how the estimation of the operational costs is a critical issue in decision making related to the type of system to adopt for the waste collection.

Keywords: decision support systems, vehicle routing, heuristics

1 Introduction

Public administrations have, during the last decades, devoted an increasing level of attention to the waste disposal problem, because of the impact on the public opinion's concern about the environment. The quantity and the different types of waste produced in the industrialized society are such that landfills cannot suffice any more to control the problem. One of the easiest ways to dispose of waste is to burn it. Burning waste allows at the same time to get rid of it, to produce energy and avoid toxic infiltration in the soil. On the other hand, it also generates among the citizens a big concern about the fumes. As a consequence, incinerators are placed as far as possible from inhabited centers and serve as many cities as possible. Recycling seems to be the most friendly way to dispose of waste: glass, paper, aluminium, green and humid can be successfully recycled in different forms. Recycling requires much effort in separation of waste, but a good level of separation reflects indirectly also upon the efficiency of the incinerators - for instance, separation allows to divert inert materials to landfill and humid waste to plants which produce compost.

In general, the community has to pay a cost for every ton of waste produced. The disposal cost per ton is affected by the "quality" of the waste: an incinerator will ask a higher cost if the waste contains inert materials or

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

humid waste, recycling plants will also price the waste upon its quality. Good quality of waste can be attained by separation. The best rule for separation is: never mix the different types of waste – which implies a change of mentality and habits for the citizens and an increment of collection costs for the community. Let us note that recycling plants, as well as incinerators, may be quite far away from the collection area and the transportation costs are themselves a relevant entry in the economic evaluation of the waste disposal system.

In conclusion, the disposal of waste involves many social and economic factors. An economic evaluation of a disposal system must take into account three main cost entries: collection, transportation and disposal.

According to what we pointed out above, choosing between differentiated and undifferentiated collection is quite a critical decision. In general, differentiated collection increases collection costs but leads to a decreasing of the disposal costs. The specific system adopted to run the service affects collection and transportation costs.

A first system – say *traditional* – was typical in earlier times and is still in use in most of the cases nowadays: a truck with a three men crew drives from one house to another and two men empty the dustbins or throw the dust bags into the rear of the truck. This system has two main shortcomings: it requires a high number of men to do the collection and, in the case citizens periodically put garbage bags on the street, imposes to fix a collection schedule for the same neighborhood. Some municipalities imposed the use of special dustbins instead of bags and reduced the crew to the single driver who at each dustbin gets off the truck, moves the dustbin close to the rear end of the truck and operates on a semi automatic system which empties the dustbin. The personnel is reduced but the time spent on the same dustbin is substantially increased.

The second system, which we call *side–loader system*, requires the citizens to throw the garbage into big street containers and is operated by a truck which stops beside a container and, by means of a semi automatic system controlled by the driver from the inside, lifts and empties the container into the truck body. This system requires the least number of operators, collects more quantity in any single operation with respect to the traditional system, is faster and does not impose to partition the collection area in neighborhoods.

The third system, which we call *side–loader system with demountable body* – operates in a quite similar way to the second one, but the trucks can unload a full body and load an empty one. The advantage is that the collection phase can be separated from the transportation phase. Specialized (and expensive) trucks can be entirely devoted to the collection while transportation is run by common trucks. This system has a smaller capacity of the body and a higher cost with respect to the system with fixed body.

When decision makers have to choose among different systems they have to take into account many variables. Among these variables, the cost plays a very relevant role. Thus, in particular, they have to estimate both the

investment and the operational costs. While standard and simple methods are well known for the investment costs, the estimation of the operational costs is much more difficult. Thus, a tool for the estimation of the operational costs would be of great help.

In this paper we propose a unique model for the estimation of the operational costs of each of the three different collection systems mentioned above. Moreover, we discuss the application of the model to two real case problems.

The paper is organized as follows. In the rest of this section we present the literature on waste collection problems. First we give a description of the waste collection problem we deal in this paper. Then we recall the features of the model presented in Angelelli and Speranza (2001) and describe how it can be used to model each of the different collection problems generated by the three different collection systems. Afterwards a sketch of the algorithm adopted to solve the model is given. In the last two sections we present two case studies worked out in the framework of the LIFE project SELECTIVE supported by the European Union. The case studies concern the mountain region of Val Trompia (Brescia - Italy) and the urban area of the municipality of Antwerp (Belgium). Finally, we draw the conclusions.

1.1 Literature on Waste Management

A number of works on waste management are available in the literature. Some attack the operational costs, such as the paper by Beltrami and Bodin (1974) where the routing costs are minimized. Others take into account multi criteria objectives or discuss the utilization of decision-aid tools for more general waste management problems where economic and social aspects are involved. Bloemhof-Ruwaard et al. (1996) formulate a variant of the two level capacitated location problem to minimize fixed and variable costs where the location of capacitated waste disposal plants and the flow of waste are studied. In Eisenstein and Iyer (1997) a case study of waste collection in Chicago is studied, where the block structure of wards simplifies the routing part of the problem, and a Markov decision model is presented to schedule the visits of trucks to the dumpsite in such a way that the capacity utilization is increased. The case where intermediate facilities are used for temporary deposit of waste before transshipment to final destination is studied by Rahman and Kuby (1995), where the trade-off between the reduction of costs and public opposition to such facilities is examined. Chang et al. (1997) formulate a multi-objective model for the minimization of the total collection distance, the additional collection costs and the total collection time, and show how a geographic information system (GIS) can help the management in analyzing different scenarios before taking decisions. Finally, Hokkanen and Salminen (1997) report on utilization of ELECTRE III – a multicriteria decision-aid tool – for the management of solid waste in presence of economic, environmental and political objectives.

2 Description of the Waste Collection Problem

Differentiated waste collection presents some features which a manager must cope with independently of the system adopted to run the service. In the typical case a fixed number of vehicles are available to run the collection which is planned on a weekly basis. A number of collection points, spread over the collection area, are given and every point must be served with a given frequency. This means that if the frequency for a collection point is twice a week, the manager can choose to serve the point either on Monday–Thursday or on Tuesday–Friday or on Wednesday–Saturday and in this case {(M–Th.), (Tu–F), (W–S)} is a list of so called *feasible visiting schedules* for the point. While the service at each collection point must be frequent enough to prevent any space or hygienic inconvenience to the citizens, it seems sensible to say that a very frequent service would increase the routing costs with no relevant advantage for the citizens.

Every day all or part of the available vehicles exit from the depot and start a collection route (dedicated to a specific type of waste). When the body is full a vehicle moves to some change over point (COP) which according to the particular system adopted can be a landfill, a disposal plant or just a "warehouse" where it deposits the body, loads an empty one and either starts a new collection route (possibly with another type of waste) or goes back to the depot if the shift is over. In the case of the warehouse, the manager will plan the transportation of the waste to its final destination by other means. In general, there can be a number of COPs, especially in differentiated collection, since paper may be carried to a recycling plant, the green fraction (GFT) to a factory and the rest to an incinerator. If the collection is based on a demountable body system the availability of more than a single COP may help to reduce the time spent in the collection area.

The objective of the manager is to minimize the total length of the routes while satisfying the constraints on the vehicle capacity and the duration of each daily route.

It is natural to model the problem as a generalization of the Vehicle Routing Problem (VRP) (a recent survey on the vast literature about VRP and its variants can be found in Laporte (1997)). The VRP usually considers a fleet of capacitated vehicles which leave the depot, serve a number of clients and go back to the depot. The waste collection problem cannot be modelled as a VRP for various reasons. Let us briefly see these reasons. First, the VRP models the case where a vehicle returns to the depot immediately after its first visit to a COP while in general a vehicle can start the collection again up to the end of the shift. Secondly, the optimization is made on the routes of a single day while the waste collection problem requires optimization over the whole period and involves the assignment of a collection schedule to each collection point. The latter issue is faced by the Periodic Vehicle Routing Problem (PVRP), which first assigns a visiting schedule to each client (a collection point in our case), and afterwards calculates the best routes for

every truck on each day of the week (see Cordeau et al. (1997) for a recent paper on PVRP). However the PVRP provides routes with a single collection route, where a collection route is a sequence of containers emptied between two consecutive visits to the COPs. Finally, the Periodic Vehicle Routing Problem with Intermediate Facilities attacks the case where more collection routes are allowed (see Angelelli and Speranza (2001)).

3 The Model

In this section we present the general model into which the various waste collection problems, generated by the different systems, fit. We first present the model and then show how the various problems fit into it.

A number T of days, the duration D of the daily shift, and a fleet of m vehicles with capacity Q_h, $h = 1, ..., m$ are given. Daily routes for the vehicles must be designed on a graph $G = (V, A)$ where $V = \{v_0, v_1, ...v_k, v_{k+1}, ...v_{k+n}\}$ is the set of the vertices and $A \subset \{(i,j) \mid i,j \in V\}$ is the set of the oriented arcs which make it possible to go directly from vertex i to vertex j. More precisely, the vertex v_0 represents the depot, vertices $\{v_1, ...v_k\}$ represent the set of k COPs and vertices $\{v_{k+1}, ...v_{k+n}\}$ represent n collection points. Not all the pairs (i,j) are included in the set A. In particular, no arc goes from the depot to any one of the COPs, since when a vehicle exits from the depot it has no load to dump at a COP. For the same reason, there is no arc connecting pairs of COPs. Besides, no arc goes from a collection point to the depot because we do not want any vehicle to finish its shift before it has dumped its load at a COP.

Every vertex i is given a service time t_i which is the time spent by a vehicle every time it visits the vertex. The service time at the depot is equal to zero. Collection points are given a list of *feasible collection schedules*. Every collection schedule defines the demand (the amount of waste to collect) of the collection point on each one of the T-day horizon (e.g. in a 5-day horizon the collection schedule {0, 10, 0, 5, 0} tells us that on days 1, 3 and 5 the demand is zero and no visit is required, while a visit is required on days 2 and 4 with a demand of 10 and 5, respectively). Note that a collection schedule is a visiting schedule with the additional information of the demand on each day of visit.

Every arc (i,j) is given a length d_{ij} and a time t_{ij} which represent the space distance and the traveling time from vertex i to vertex j, respectively.

We call *route* a circuit on the graph G which goes through the depot and *collection route* any sub-path of a route which starts from a COP (or the depot), goes through a number of collection points and finishes at another COP (possibly the same).

A solution of the model is a set of routes assigned to each vehicle on each day of the horizon such that every collection point is visited according to one of its feasible collection schedules and no more than once on the same day.

Note that no constraint is given on the number of visits a COP can receive on any day, that is many vehicles can visit many times the same COP or a COP can be not visited at all during the horizon.

A solution is feasible if all the routes satisfy the duration bound D and the capacity Q_h of vehicle h is not exceeded by the demand of the collection points visited on any of the collection routes. We want to find a feasible solution which minimizes the overall distance traveled by the vehicles. This problem is the PVRP with intermediate facilities introduced in Angelelli and Speranza (2001) where the COPs play the role of the intermediate facilities.

The model lends itself to different contexts such as differentiated and undifferentiated collection, the traditional and the side–loader systems with or without demountable body. The application of the model to any of the three collection systems for undifferentiated collection is straightforward: the collection points represent dustbins or street containers for traditional or side–loader systems, respectively.

In the case of differentiated collection a vehicle cannot mix different types of waste. Thus, we need to modify the arcs of the graph which connect containers of different type. We can either delete the arcs from the graph or equivalently set the travelling time and distance attributes to infinity so that no feasible solution can contain those arcs. Furthermore, if a COP represents, for instance, a recycling plant for paper it cannot accept any other type of waste. In such cases we delete all the arcs which go from a container to any incompatible COP. There is, in general, no restriction on the arcs from a COP to the containers. There is another important issue in applying the model to differentiated collection. The capacity of the vehicles may strongly depend on the type of waste they collect. Since in the model such capacity is fixed, we should think of it as a virtual capacity and multiply the real demand of each container by the ratio between the virtual capacity and the real capacity for the type of waste collected. Such operation ensures that the virtual capacity is violated by virtual demand if and only if the real capacity is violated by real demand.

4 The Solution of the Model Applied to Real Instances

4.1 Application of the Model to Real Instances

The described model assumes data to be available for the calculation. Actually a large part of this data is not directly available in real problems for different reasons. Many are just not deterministic: the travelling time between two vertices, the service time at COPs, the service time and the demand at collection points, are random variables. Since our model is deterministic we use an estimation of the mean value of any random variable in the model.

Other data may be not available because of the tremendous effort it would be necessary for its estimation (e.g. distances between each pair of a large number of collection points). In our two case studies, the problems contained

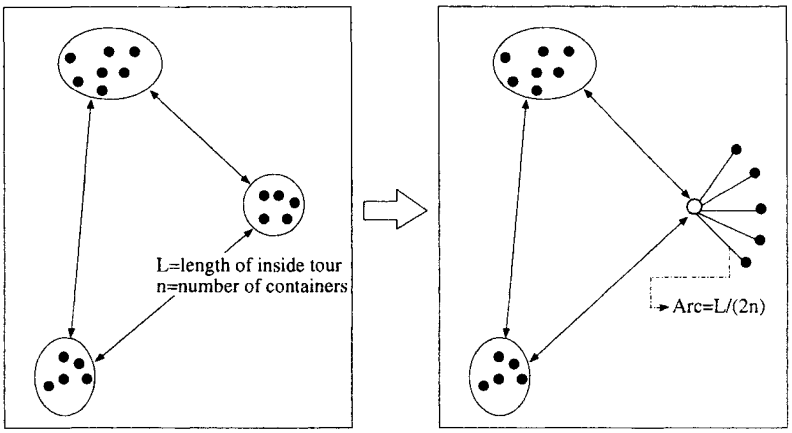


Fig. 1. Modeling by macro-points

from 181 to 345 vertices which implies more than 16,000 distances in the most favorable case and since no electronic representation of the road network was available there was no practical way to calculate the exact distance between each pair of vertices of the model. A compromise was found when the managers were able to provide aggregate information on sets of collection points.

We found it convenient to use the concept of macro-point as an aggregation of identical collection points which are close to each other, collect the same type of waste, have the same daily accretion rate of waste and share the same visiting schedules. These points are almost indistinguishable from each other. Note that a visiting schedule together with a daily accretion rate of waste suffice to define a collection schedule. The demand on each day is calculated according to the daily accretion rate and the number of days elapsed from the previous visit (e.g. for a (M-Th) visiting schedule and accretion rate 2, the demand on Thursday is 2×3 and the demand on Monday - of the following week - is 2×4).

If the user of the model can describe the data in terms of macro-points, providing each macro-point with the number of collection points, a list of feasible visiting schedules, the accretion rate of waste and the total distance a vehicle covers to visit all the points, then it is easy to automatically rewrite such data in terms of collection points. Every macro-point is replaced by a cluster of collection points connected to a "central gate" which is the only way to reach and leave a single collection point (see Figure 1). The total distance inside the macro-point is equally distributed on the arcs of this special cluster. If the distances between pairs of macro-points are known, the distance between each pair of collection points can be calculated and the graph G can be built.

Finally, the average speed between macro-points and the average speed inside all macro-points were considered adequate to evaluate the traveling time between any pair of vertices in the graph G.

Summing up the above remarks, we report below the data structure adopted for the real instances:

- number of identical vehicles;
- duration of the shift;
- set of distances between every pair of macro-points, COPs and depot;
- average speed outside the macro-points;
- average speed inside the macro-points;
- average service time at the containers;
- capacity of the vehicles for each type of waste collected;
- description of COPs:
  - list of accepted types of waste;
  - average service time;
- description of macro-points:
  - type of waste collected;
  - number of containers;
  - distance required to visit all the containers;
  - average daily accretion rate of waste;
  - list of the feasible visiting schedules.

4.2 The Solution Algorithm

We implemented a Tabu Search algorithm (TS) to solve the problem (see Aarts and Lenstra (1997) for references on tabu search algorithms). Here we give a sketch of the algorithm which has been presented in Angelelli and Speranza (2001) for the PVRP with intermediate facilities, then describe the variants we adopted to solve the case study instances. A few changes in the base algorithm were in fact appropriate to adapt the algorithm to the macro-point structure of the instances.

First, a visiting schedule is randomly chosen for every macro-point. All the containers in each macro-point are assigned the selected schedule so that for every day the set of containers to be visited and their demand are fixed. Then an initial solution is built. A solution is a set of routes assigned to the fleet on every day so that each container is visited according to its schedule. A solution is feasible if all the routes satisfy the time and capacity constraints, otherwise it is called infeasible.

After the initial solution is created, the algorithm starts to iteratively move from a solution to another by means of "small" changes of the current solution – such changes are also called moves. After a fixed number of iterations the best feasible solution found is returned.

The algorithm considers four types of move:

- a container is removed from a collection route and inserted into another one on the same day;
- a container is assigned a different collection schedule;
- two collection routes are interrupted and differently linked again. The aim of this move is to eliminate expensive intersections between collection routes;
- the containers of two collection routes are redistributed. The aim of this move is to solve at once intricate intersections of collection routes.

The new current solution is selected among the candidates, obtained by applying the different moves in different ways, as the one which is "most promising" and not likely to have already been examined. The most promising solution is the one which minimizes a penalized cost function. Such function takes into account the value of the objective function and a measure of the violation of the time and capacity constraints. As a consequence, an infeasible solution with low value of the objective function and not too big a violation of the constraints can be selected instead of a feasible one. In order to avoid cycling on the same solutions, a list of forbidden solutions is updated at each iteration. Whichever move is chosen by the algorithm, a number of containers are moved from one collection route to others. A tabu list keeps record of such movements so that the algorithm will be able to prevent itself from doing the reverse move later. We note that for reasons of efficiency a limited number ($O(\ln n)$ where n is the number of containers) of moves are stored in the list – instead of an unlimited number of solutions – and the algorithm is guaranteed from short term cycles only.

To solve the Val Trompia case we forced the algorithm to initially assign the same visiting schedule to all the containers in the same macro-point and modified the tabu list as follows. When a container is extracted from a collection route, the tabu list records the macro-point to which the container belongs instead of the container itself. By means of this device, the algorithm avoids a useless exchange, between two collection routes, of containers which belong to the same macro-point and are perfectly equivalent to each other.

Differently from the Val Trompia case, a macro-point in the Antwerp case is composed of a small number of containers. This suggested the idea of an algorithm which moves the containers within a macro-point all at once as if they are a single big container. We tested such a variant versus the basic algorithm used for the Val Trompia case on the Antwerp case. As the modified algorithm turned out to be faster and more effective we adopted it for all the simulations.

5 The Val Trompia Case

The Val Trompia is a valley in the province of Brescia, Italy, where an undifferentiated collection of waste is run by a private company. (See road map in

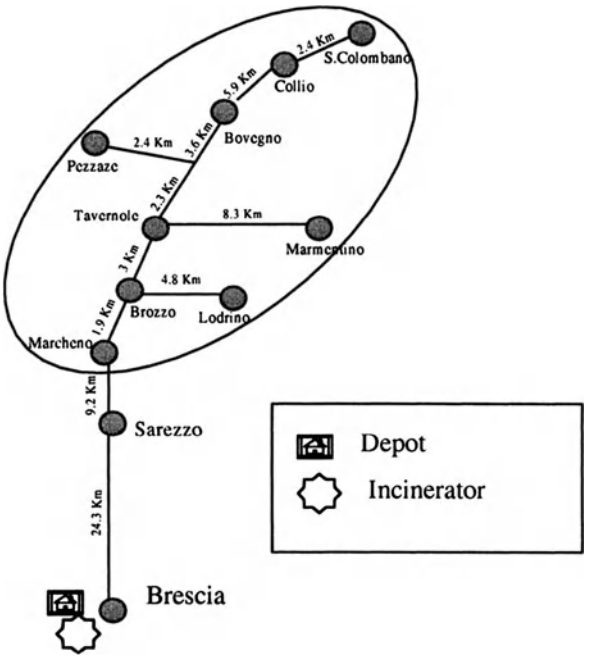


Fig. 2. Road map of Val Trompia

Figure 2.) The collection is planned on a weekly basis for an average amount of 117 tons of undifferentiated waste. The waste is collected from 343 street containers distributed among 10 villages. The company is engaged, by an agreement with the municipalities, to do the collection twice a week in all the villages except Sarezzo which is the most populated one and requires the collection trice a week. Collection schedules are not fixed by the municipalities, but can be rather freely chosen by the company for each container. Nevertheless, the days of such collection schedules are supposed to be uniformly distributed over the week (see Table 1 for the list of feasible collection schedules in Val Trompia). We present the situation at the time of the case study. The collection is performed by 2 side-loader vehicles, called for short CMPLs, which work every day, but Sunday, 6 hours a day. The load capacity of the CMPL is 13.0 tons. The service time at the containers and at the incinerator is 55 seconds and 10 minutes, respectively. The average speed between and inside villages is 30 Km/h and 23 Km/h, respectively. The waste is delivered to an incinerator located in Brescia very close to the depot. A part of the daily route of the vehicles is used for the transportation of the empty body from the depot to Sarezzo which is located at the beginning of the valley and is on the way of all the vehicles which visit any other village of the valley. When the body of a CMPL is filled up the same vehicle has to transport the waste to the incinerator and covers at least the distance

Table 1. Val Trompia, visiting schedule codes

Code	Visiting days
21	Mo-We-Fr
37	Mo-We-Sa
42	Tu-Th-Sa
41	Mo-Th-Sa
9	Mo-Th
18	Tu-Fr
36	We-Sa
17	Mo-Fr
34	We-Sa

Table 2. Val Trompia, details of villages

Village	Number of containers	Tons of waste collected per week	Inside distance (Km)	Feasible visiting schedules
SAREZZO	158	60	47.0	21 37 42 41
MARCHENO	46	14	12.0	9 18 36 17 34
BROZZO	15	5	1.7	9 18 36 17 34
LODRINO	15	4	4.0	9 18 36 17 34
TAVERNOLE	16	5	1.8	9 18 36 17 34
MARMENTINO	11	4	1.2	9 18 36 17 34
PEZZAZE	20	5	2.3	9 18 36 17 34
BOVEGNO	29	10	10.9	9 18 36 17 34
COLLIO	21	7	7.9	9 18 36 17 34
S.COLOMBANO	12	4	4.5	9 18 36 17 34

between Sarezzo and Brescia. The time spent for transportation (of empty and filled bodies) prevents the vehicles from a second round trip on the same day. (See Table 2 for details on the collection in each village.)

The company which runs the service has to decide whether to move from the present side-loader system without demountable body to a system with demountable body. A change of system would not imply any change for the citizens. Thus, the economic evaluation of the two systems represents the critical issue for the decision. Given the relatively long life of the vehicles, the operational cost is the critical part of the economic evaluation.

5.1 Simulations

Our aim is to compare the operative costs of the CMPL system with the side-loader system with demountable body, called for short CWS, when one

Table 3. Val Trompia, results of simulations

Number of Vehicles	Capacity	Duration of Shift	Depot	COP	Distance for collection	Distance for transportation	Total distance
2 CMPL	tons 13.0	h 6.00	Brescia	–	Km 1,102	–	Km 1,102
2 CMPL	tons 13.0	h 6.00	Brescia	–	Km 1,059	–	Km 1,059
1 CMPL	tons 13.0	h 9.00	Brescia	–	Infeasible	–	–
1 CWS + 1 truck trailer	tons 9.5	h 6.00	Sarezzo	Sarezzo	Km 582	Km 340	Km 922
1 CWS + 1 truck trailer	tons 9.5	h 8.00	Brescia	Sarezzo+ Marcheno	Km 855	Km 360	Km 1,215
1 CWS + 1 truck trailer	tons 9.5	h 9.00	Brescia	Sarezzo	Km 874	Km 340	Km 1,214
1 CWS + 1 truck trailer	tons 9.6	h 9.00	Brescia	Sarezzo	Km 872	Km 340	Km 1,212
1 CWS + 1 truck trailer	tons 10.5	h 9.00	Brescia	Sarezzo	Km 863	Km 340	Km 1,203
1 CWS + 1 truck trailer	tons 10.6	h 9.00	Brescia	Sarezzo	Km 837	Km 340	Km 1,177

or two COPs are placed within the collection area, only 1 CWS is used for the collection and 1 truck trailer is used for the transportation of the waste from the COP (or COPs) to the incinerator plant. Different assumptions are made on the number and the location of the COPs and on the location of the depot. We also exploit the impact of an increase of the duration of the daily shift from 6 to 9 hours, in view of a possible agreement between the private company which runs the service and the Trade Unions. Moreover, we test the robustness of the cost estimations with respect to the capacity of the CWS which can be affected both by technological improvements of the vehicle and by variations of the composition of the waste collected.

In Table 3 we report the results on the current CMPL system and the CWS system under the different assumptions. In the first row of the table the solution adopted in the actual situation is summarized. In the second and third rows we present the results obtained by running the tabu search algorithm with the 2 actual CMPLs and the actual 6 hours shift and with 1 CMPL and a 9 hours shift. The latter simulation has been run to check the feasibility of a solution with 1 CMPL only. We note that the current routing can be slightly improved, but the service cannot be performed by a single CMPL vehicle even if a 9 hours daily shift is allowed. In rows 4-9 we present the results of the various simulations run for the CWS system. A reduction of the total distance (shown in the last column) can be attained with a single CWS if both the depot and a single COP are located in Sarezzo. Note that in this case a 6 hours shift is enough to run the service. In the present case where the depot is located in Brescia 8 hours are needed if we place 2 COPs in Sarezzo and Marcheno, while the maximum duration allowed for the shift is required if we use only one COP in Sarezzo. In the latter part of the table we report the sensitivity analysis on the capacity of the CWS under the assumption of a single COP located in Sarezzo. As expected, the distances

decrease as the capacity increases, although the variations of the distances are not proportional to the variations of the capacity.

The solution with 1 CWS, and two COPs in Sarezzo and Marcheno, turns out to be the most convenient one, when the investment costs and the personnel costs are considered. This is mainly due to the fact that this solution involves 1 expensive CWS and 1 common truck trailer while 2 expensive CMPLs are needed in the present situation. In conclusion, the company which runs the collection service in Val Trompia has decided to move to the CWS system. The system is now fully operating providing an increased level of the service with cost reduction.

6 The Antwerp Case

The company which runs the collection service in Antwerp, Belgium, has adopted a CWS system to run a differentiated collection of paper and GFT on a selected set of condominiums spread over the urban area. The collection is planned on a weekly basis and at the very first stage of the implementation it involved an average amount of 14.5 and 16.5 tons of paper and GFT, respectively, per week from 107 paper-containers and 72 GFT-containers. The containers are geographically concentrated in small groups, which become the macro-points, located where groups of condominiums are located. The number of such macro-points is 42 and 26 for paper and GFT, respectively. A list of feasible visiting schedules for the containers is fixed by the company, according to the rate of accretion at each macro-point. (See Table 4 for a list of possible collection schedules in Antwerp.) The collection is performed by 2 CWS on Monday, Tuesday, Thursday and Friday. Wednesday is dedicated to maintenance. The load capacity of the CWS is 5 and 11 tons for paper and GFT, respectively. Paper is collected on each operative day while GFT is collected on Monday and Tuesday only. The daily shift is 6 hours. Two COPs are available in the city for the exchange operation, which takes 10 minutes, of a filled body with an empty one. Paper is delivered to a recycling plant about 20 Km North of Antwerp while the GFT is sent about 30 Km South to a plant which produces compost.

6.1 Treatment of the Available Data

The data on the Antwerp instance were provided by the company which runs the service in the following form.

Each macro-point is identified on a map of the city by means of a small circle which covers the small area where the containers are located. Similarly to the macro-points, the COPs and the depot are identified on the map by small circles.

The data of the problem were made available through the daily collection tables of a "typical" week. The tables, available for each collection vehicle, reported the following information:

Table 4. Antwerp, visiting schedule codes

Code	Visiting schedules for paper	Visiting schedules for GFT
1	Mo	Mo
2	Tu	Tu
8	Th	
16	Fr	
9	Mo-Th	
18	Tu-Fr	

- day of collection and type of waste collected;
- departure time from the depot and mileage;
- for every macro-point visited: number of containers, quantity of waste collected, arrival time and mileage, departure time and mileage;
- total weight of the waste collected and mileage at the COP;
- mileage at the depot in the evening.

The quantity of waste reported in the tables was expressed in number of quarters of a container.

Some of the data required by the instance data-structure are directly available in the tables provided by the company. Other data have to be extrapolated from the tables:

- the service time at the containers is estimated as the average time spent in a macro-point with one container only;
- the average traveling speed between macro-points is deduced from the arrival and departure mileage and time;
- the average weight of a quarter of a container is estimated for each type of waste by the ratio between the total weight of waste and the number of quarters collected during the week;
- the daily accretion rate in a macro-point is calculated as the total number of quarters collected during the week multiplied by their average weight and divided by 7;
- the feasible visiting schedules are computed as follows. First, the minimum number of visits is calculated such that overflow of waste is avoided (e.g. if 5 quarters are collected from every container in a macro-point during the week, then at least two visits are necessary). The result is that all the GFT containers may be serviced only once a week and two feasible visiting schedules are generated for all GFT containers: the first schedule requires only one visit on Monday while the second requires only one visit on Tuesday. Obviously, the demand of the container on the visit day is equal to the total waste collected in the week. Some of the paper containers must be visited at least twice a week, others can be visited only once. Two feasible visiting schedules are generated for all paper containers: the first requires visits on

Monday and Thursday, while the second requires visits on Tuesday and Friday. Four further visiting schedules of only one visit are added to the list of those paper containers which do not need two visits a week;

– the average traveling speed inside the macro-points is computed using the distances as they are reported in the collection tables and the corresponding time diminished by the average service time spent at the containers;

– the distances between macro-points, COPs and depot are calculated as follows. First, we assign to every macro-point, as well as to the depot and the COPs, a pair of coordinates drawn from the map of the city so that the Euclidean distance between any pair is easily calculated. Then, we calculate a transformation coefficient as the average ratio between true distances (reported in the collection tables) and the corresponding Euclidean distances on the map. Finally, we use as "real" distances the Euclidean distances multiplied by the transformation coefficient.

6.2 Simulations

The company which runs the collection service in Antwerp has only recently moved from the traditional system to the CWS system which is at present in an experimental phase. The company has to decide whether it is worthy to extend the CWS system to the rest of the city or to limit its application to the condominiums. The company believes that the cost of the CWS system, measured in cost per ton of collected waste, is very high, but that this may due to the fact that the capacity of the vehicles is not completely used and that the vehicles spend most of the time traveling because the condominiums are spread over the urban area. The problem is to estimate the operational costs when the capacity is fully used and when the containers are closer to each other.

In order to give a proper answer to the estimation problem, we have simulated some scenarios. In the "Augmented" scenario the production of waste is increased so that each CWS can be used for two full load collections per day. In the "Shrunk 1" scenario the distances between macro-points are halved and the production of waste is increased so that we can investigate the case where the CWS vehicles are used more intensively for collection instead than for traveling between very sparse collection points. In the "shrunk 2" scenario the logic of the "shrunk 1" scenario is repeated and the distances are halved again.

In Table 5 we report the results of the simulations run on the Antwerp case. In the "Collection" column we report the number of containers, the quantity of waste collected in tons, the distance traveled by the CWSs and the number of bodies unloaded at the COPs during the week. In the "Transportation" column the number of trips to the disposal plants and the distance traveled by the truck trailer are reported. Finally, in the "Totals" column the total distance traveled and the distance per collected ton are reported. The

Table 5. Antwerp, results of simulations

Collection				Transportation		Totals		
containers	weight	distance	bodies	trips	distance	distance	Km/ton	
								Initial
107	tons 14		4	2	Km 80			Paper
71	tons 17		2	1	Km 70			GFT
178	tons 31	Km 253	6		Km 150	Km 403	**13.0**	**Total**
								Undifferentiated
178	tons 31	Km 207	5	3	Km 165	Km 372	**12.0**	
								Augmented
134	tons 23		6	3	Km 120			Paper
74	tons 43		5	3	Km 210			GFT
208	tons 66	Km 385	11		Km 330	Km 715	**10.8**	**Total**
								Shrunk 1
155	tons 28		7	4	Km 160			Paper
85	tons 50		7	4	Km 280			GFT
240	tons 78	Km 275	14		Km 440	Km 715	**9.2**	**Total**
								Shrunk 2
168	tons 29		7	4	Km 160			Paper
124	tons 69		6	3	Km 210			GFT
292	tons 98	Km 196	13		Km 370	Km 566	**5.8**	**Total**

distance traveled per collected ton allows us to compare the different situations and can be easily transformed in cost per ton to obtain the operational cost per unit of collected waste, a parameter which is of immediate impact and interpretation for the company. In Table 5 we also show the "Initial" scenario which corresponds to the real distances and quantities of waste and the "Undifferentiated" scenario which is the initial scenario where no distinction is made between paper and GFT containers (in a single collection tour, both paper and GFT containers may be emptied). The latter simulation is carried out to estimate the cost of the differentiated collection with respect to the undifferentiated one.

As we may expect, when an undifferentiated collection is performed the total distance traveled decreases as well as the distance per ton. However, even limiting the analysis to the economic aspect of the problem and ignoring the relevant environmental consequences, the disposal costs in the differentiated collection are much lower, since no disposal cost is paid for the paper and a low cost is paid for the GFT disposal.

The other scenarios show how the performance ratio, distance per ton, improves when the capacity if fully used and much more when the capacity is fully used and the distances to be traveled are reduced. A more intensive use of the CWS system on the urban area would generate such a situation. The Municipality is now increasing the number of vehicles involved in the collection.

7 Conclusions

The economic evaluation of a waste collection system requires an estimation of the operational costs. In this paper we presented a model to estimate the operational costs of a waste collection system which are strictly related to the distance traveled to collect the waste and deliver it to the disposal plants. The model can be applied to different collection systems (traditional, side–loader, side–loader with demountable body) so that a comparison between the systems can be performed. We applied the model to two case studies. In the first case we compared the operational costs of a side-loader system versus a side–loader with demountable body under different scenarios on the capacity of the trucks, on the duration of the shift and on the location of the COPs. In the second case we used the model to estimate the operational costs of a side–loader with demountable body system under different scenarios: undifferentiated and differentiated waste collection, extension of the collection area and quantity of waste to be collected. Practical issues such as the management and treatment of the available data in real problems are also addressed.

References

Aarts, E.H.L. / Lenstra, J.K. (eds.) (1997): "Local Search in Combinatorial Optimization" John Wiley & Sons Ltd.

Angelelli, E. / Speranza, M.G. (2001): The Periodic Vehicle Routing Problem with Intermediate Facilities. *To appear on Eur J Opl Res.*

Beltrami, E.J. / Bodin, L.D. (1974): Networks and vehicle routing for municipal waste collection. *Networks* 4:65-94.

Bloemhof-Ruwaard, J.M. / Salomon, M. / Van Wassenhove, L.N. (1996): The capacitated distribution and waste disposal problem. *Eur J Opl. Res* 88:490-503.

Chang, N.B. / Lu, H.Y. / Wei, Y.L. (1997): GIS Technology for Vehicle Routing and Scheduling in Solid Waste Collection System. *Journal of Environmental Engineering* 123:901-910.

Cordeau, J.F. / Gendreau, M. / Laporte, G. (1997): A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks* 30:105-119.

Eisenstein, D.D. / Iyer, A.V. (1997): Garbage Collection in Chicago: A Dynamic Scheduling Model. *Management Science* 43:922-933.

Hokkanen, J. / Salminen, P. (1997): Choosing a solid waste management system using multicriteria decision analysis. *Eur J Opl. Res* 98:19-36.

Laporte, G. (1997): Vehicle Routing. In Dell'Amico M, Maffioli F, Martello S (eds). *Annotated Bibliographies in combinatorial Optimization 1995.* John Wiley & Sons Ltd, pp. 223–240.

Rahman, M. / Kuby, M. (1995): A multiobjective model for locating solid waste transfer facilities using an empirical opposition function. *INFOR* 33:34-49.

Strategic Vehicle Routing in Practice – A pure Software Problem or a Problem Requiring Scientific Advice? Routing Problems of Daily Deliveries to the Same Customers

Roland Dillmann

Department of Economics and Business Administration, University of Wuppertal, 42097 Wuppertal, Germany

Abstract. For improving their delivery tours, newspaper and magazine wholesalers require a practice-oriented solution to the strategic vehicle routing problem. The main tool to achieve this goal is a special software but if the most is to be gained from the existing optimization potential, it is necessary to have a detailed knowledge of the problem in its real context. This knowledge can be easily acquired by solving the problem in dialog. This paper discusses a practicable approach to the press distribution problem. It presents the situations where software tools are helpful and how they should be applied. Furthermore, a working-in-dialog procedure is described.

1 Introduction

We are faced with many difficulties when trying to solve the problems of strategic vehicle routing in practice. By planning the press distribution process for nearly 35 out of the 106 wholesalers located in Germany, we obtained the practical experience at the background of this paper. It is structured as follows: First of all, we will give a brief description of the various aspects of the problem. Then we will discuss the data issue and offer our proposal as to which data should be introduced into the solution process. After that, we will demonstrate how software can be used during the solution process, and why it is necessary to integrate a dialog-based approach. A final chapter gives a summary of successful and failed projects.

2 Description of the Problem

At the moment, there are 106 wholesalers in Germany. They supply 120,000 customers every day. However, the situation is changing. We can observe a concentration of wholesalers due to mergers, so that the number of wholesalers does obviously vary. Furthermore, due to the fact that the publishers of the newspaper

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

"Bild" currently intend to introduce a later deadline, we expect wholesalers to increase the number of their depots.

The press distribution problem is a strategic vehicle routing problem with time windows and restricted loading capacities. Every night the publishers of newspapers and magazines deliver their products exclusively to wholesalers. They have to organize daily deliveries in such a way that they meet the individual deadlines fixed for every retailer. We advised nearly 35 wholesalers on how to solve the daily press distribution problem. The number of their customers, i.e. retailers, varied between 400 and 2,700, and the number of tours between 14 and 67. An average tour serves more than 40 customers. Usually, vehicles with a loading capacity of 1.2 to 2.4 tons are used.

What is the situation we are confronted with? Although every customer has a latest delivery time, only very few customers have an earliest delivery time, as it is in most cases possible to lock the consignment up in special boxes for which the drivers have the keys. Therefore, the customer is usually not present during delivery. Only when drivers cannot enter certain locations with special traffic regulations early in the morning, e.g. health-resorts, or have to observe ramp times, e.g. in case of supermarkets, we are restricted by two-sided time windows.

The volume of the delivery is subject to seasonal and daily fluctuations because wholesalers have to distribute a variety of publications. The final version of a daily newspaper is printed around midnight and delivered late in the early morning hours. As it is not possible to start delivery earlier, the distribution of daily newspapers presents one of the most time-restricted delivery problems. In regions where different newspapers of regional importance are published daily, the time-window problem is often dealt with in such a way that the consignment is picked up by drivers at a special location and then put together during the tour. This means that, when supplying the customer, the driver takes the number of copies ordered from the consignment he has picked up and adds these copies to the package, which has been prepared at the wholesaler's warehouse. Due to this additional work, the duration of these tours is longer. On the other hand, they can start earlier.

Many magazines are printed weekly, fortnightly, monthly or irregularly. Most of these are published on Wednesdays, Thursdays or Fridays. Especially TV guides are rather heavy, and millions of copies have to be distributed. Fortnightly, they considerably increase the number of copies to be delivered. At first sight it seems to be attractive that delivery schedules change regularly according to the daily change of the number of copies to be delivered. However, this means that the tours are not stable, whereas wholesalers generally prefer stable tours. They fear problems with regard to box keys and longer delivery times.

Wholesalers try to solve the problem caused by the daily change of the required capacities by planning pre-tours for Wednesdays, Thursdays and Fridays, i.e. additional tours starting before the usual tour. Nevertheless, they are faced with the problem that they overload their vehicles on some days when planning with the capacity, which is sufficient for 40 weeks of the year. However, if they provide for capacities and times that would be sufficient for any day of the year, deliveries will become very expensive. Yet press products only yield a low trade margin.

The wholesalers' objectives apparently varied from case to case. In the course of the case study, wholesalers stated the following objectives:

- *Reduction of costs.* Only in very few cases can this be achieved by reducing the travel distance, as drivers refuse to accept a reduction of work and, thus, of their wage. The best way to reduce costs is to reduce the number of tours by redistributing the work on fewer tours and then to reduce the travel distances.
- *Reduction of tours.* This is rarely stated as an explicit objective.
- *More tours in case of enormous overloading* is an objective explicitly expressed in several projects.
- *Arriving earlier at customers' who open early.* This task is stated in most cases.
- *Homogeneous tours.* Wholesalers never formulated this objective at the beginning, but they often considered it a desirable goal when discussing the new proposals for the tours. Wholesalers realized that they could sell homogeneous tours more easily. A definition of what should be regarded as homogeneous is to be made. We defined a routing plan as being homogeneous if the least time-consuming tour only lasts somewhat longer than the most time-consuming tour. A definition based on the number of customers, capacities or distances seemed to be inadequate because of the topography of the regions to be served.
  We were often faced with the situation that the current routing plans were well organized in the region near the depot, but that there were overloaded tours in case of long distances. Customers located far away from the depot were served very late. We explicitly recommended to wholesalers that tours with long distances should be tours with low capacities. It is very difficult to organize pre-tours if the main part of the consignment is to be brought to customers located far away from the depot. It should be taken into consideration that in most cases pre-tours are added to the main tour and done by the same driver.

In all cases we studied, we found that wholesalers felt that their problem required a multi-criteria decision. Therefore, route planning in practice cannot be seen as an isolated task. It is the wholesaler's job to schedule the product range of every customer, and to put the consignment together every morning. A bad organization at the time of compiling the consignment and loading the vehicles is time consuming and will increase the delivery problem, as tours cannot leave the depot as early as they should. In such cases, we suggest an organizational pre-processing. Such an organizational pre-processing includes the reorganization of the process of putting the consignment together as well as dealing with the problem of how to handle newspapers published late, or how to organize the supply of customers at a very isolated location. Possible solutions with regard to these special cases range from delivering the consignment to the customer's address early in the morning to arrangements with bus drivers to deliver the consignment to a nearby bus stop. A considerable amount of creativity should be invested to relax the real problem instead of trying to solve an unrelaxed problem mathematically.

In some cases a reorganization of the loading process was the fundamental step towards achieving a punctual delivery. A simple reallocation of machines provided for the use of more assembly lines early in the morning. In another case, drivers first put the consignment together, and then started their tour only when

the complete process was finished. We achieved an earlier time of delivery simply by hiring additional staff for loading, so that it was not necessary for the drivers to be present when their vehicles were being loaded.

3 The Data Problem

Definite data is required to solve the press distribution problem. However, what we have to measure in order to obtain the desired data is rather complex. The following two examples are to show this. First of all, there are the exact departure times. The time of departure differs from day to day by a few minutes because the time at which wholesalers receive the time-critical newspapers varies (especially in case of "Bild". "Bild" is important for two reasons: First of all, it is the best-selling daily newspaper in Germany, and second, it is exclusively distributed by wholesalers, in contrast to subscriptions; with more than 1,200,000,000 copies sold per year, "Bild" alone accounts for nearly 10% of the whole press business comprising more than 3,000 publications altogether). The newspaper wanting to present the latest news, an important sports event may cause a later deadline. Thus, due to the importance of departure times of tours, it seems to be essential to find out on which times a new routing plan should be based.

Second, there are the capacities. The loading weight of 1.2 – 2.4 tons per vehicle is not a problem of loading space but only of weight. Loading the vehicles until all loading space has been used up leads to an extreme overloading. Due to the fact that the weight of the newspapers and magazines distributed during the least busy week often accounts for less than 50% of the weight of those distributed during the busiest week, a second question should be raised: Which weeks of the year should be regarded as typical with regard to capacity utilization, and therefore should serve as a basis for defining the capacity restrictions and loading times on which all time schedules and evaluations with regard to punctuality will be based?

To find a successful and practice-oriented solution to the press distribution problem, we propose the information base described in the following. As we had to estimate the required data in many cases, we will explain our approach in detail. The person planning the tour has to be informed (if possible, by the wholesaler)

- *about the required geographic information.* Although road networks often form the basis of geographic information, and the user tries to identify the customer whose location is the nearest point within the network, it would be advantageous to subdivide the road section in question by introducing the customer's location as a new node. Thus, the directed road section is divided into two directed parts. Positioning the customers adequately within a network was a difficult task. For example, we encountered a lack of information on the house numbers of road sections. The information available is improving every day. However, we observe in practice that the exactness of road networks is overestimated. To gain experience with regard to the degree of exactness, the user should try to construct a network by means of digitization or to follow the GPS record of a trip on the computer. For technical reasons, parts showing a high

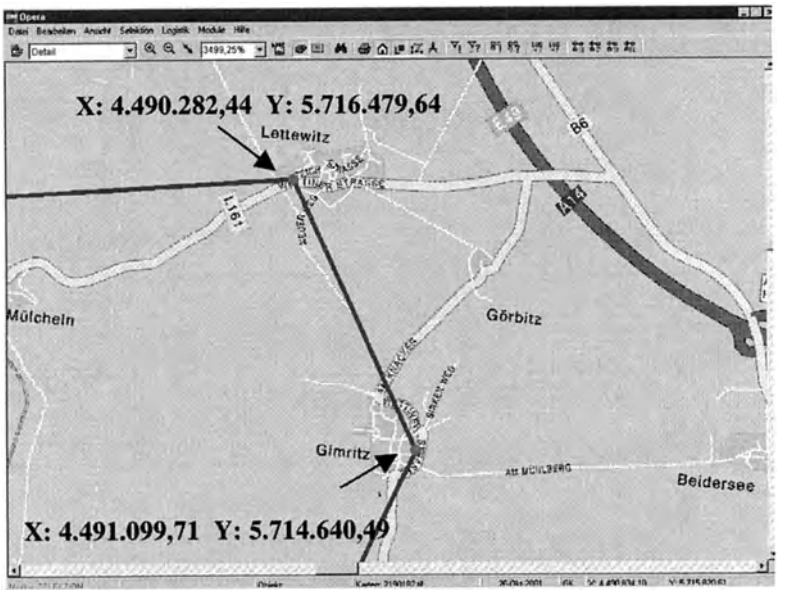


Fig. 1. Illustration of customer locations and sequences in a geographic information system

degree of exactness alternate with parts of an inadequate quality. We followed a very simple approach and used Gauss-Krüger coordinates. We digitized the customers into special maps, which assign a Gauss-Krüger coordinate to each point on the map (see Fig. 1).

As approximation of the distance between two points, we used Euclidean distances multiplied by a factor (for methods to measure distances on the basis of coordinates and estimates of the exactness of these procedures see Berens / Körling (1985), Brimberg / Love (1991), Christofides / Eilon (1969), Love / Morris (1972) and (1988), Ward / Wendell (1985), for their use in practice see Holt / Watts (1988), Stokx / Tilanus (1991)). We were very often confronted with the point of view that the use of Euclidean distances is an antiquated base for sequencing. Software companies offering application software to solve strategic problems support this opinion. They had developed this software at a time when customers had to be placed at the end points of street sections, and when networks of the main routes were available only. When trying to find a minor road in a rural region, users of such a software often have to find out much to their surprise that they can only mark their customers at some pre-selected points. Furthermore, representatives of software companies often try to convince wholesalers that routes can be taken as scheduled without encountering any problems, and some of them strictly deny the fact that sequencing in strategic situations should be considered a project requiring scientific advice.

We experienced that if the computed sequences based on Euclidean distances are checked by means of the maps they are based upon, geographical barriers will appear and can easily be avoided. As a result, our approach is sufficient for

characterizing the relative position of customers and for sequencing. But it is not suited for determining the correct distances. In fact, such a determination is only required for the final sequence (see also Fig. 5 and Fig 6.).
The data base has to be evaluated with regard to the whole procedure. The software we saw showed a mixture of Euclidean distances and network distances with an insufficient degree of digitization. For example, we tested networks in the region of the Nord-Ostsee-Kanal. We tried to cross the channel and were not informed that we had to use a ferry. As we have already pointed out, it is true that data bases are continuously improving, but they are also subject to daily changes. Euclidean distances are a good means for planning sequences when maps validate the tours and misleading information is detected. Other solutions are more convenient, but they are much more expensive and they involve a risk: They suggest that plans thus created are absolutely reliable.

Table 1: Types of customers according to opening hours

Group	Type	Opening hours
1	shops selling newspapers and magazines only	regular
2	corner shops	regular
3	food stores	regular
4	supermarkets	regular
5	hospitals, homes for the aged etc.	regular
6	gasoline stations	open 24 hours
7	kiosks	irregular

- *about the time windows within which customers are to be supplied.* This is a very important aspect, which seems to be easy to solve at first sight. But our experience has been to the contrary. To understand the difficulties involved, we subdivide the customers into seven groups (see Table 1).
  The first five groups are easy to handle. Gasoline stations, however, are often open 24 hours. They should be supplied as early as possible, as many people tend to buy a weekly when refuelling their car on their way to work. The earliest possible time of delivery depends on the time the vehicles depart from the depot. This is generally after midnight. Therefore, wholesalers have to fix a deadline for delivery to gas stations. We learned that this is a difficult task because wholesalers often have no idea of a realistic and advantageous time for serving gas stations.
  Kiosks, the last group, tend to cause further problems because there is no fixed earliest opening time required by law. Existing routing plans always included kiosks that were supplied with a delay, i.e. after opening. Many owners accept this delay and hesitate to ask the wholesaler for an earlier time of delivery. Hence, the wholesaler thinks that the customer agrees with the time he has fixed and enters this time into his list of delivery deadlines. Establishing a new distribution plan based on this data can cause serious problems. For example, if a new and improved plan is introduced to kiosks both located in the same neighborhood and opening at the same time. Before the new plan was intro-

duced, all kiosks had been supplied with a slight delay. Let us further assume that after the improvement all kiosks are supplied on time, but not at the same time. This means an important change with regard to competition among the neighboring kiosks. Usually, early customers, such as construction workers, buy some cigarettes, something for breakfast, a bottle of beer or other alcoholic drinks at the kiosk. If possible, they also want to buy a daily newspaper, especially "Bild". It goes without saying that many of these early customers will buy "Bild" at the kiosk, which opens first. As a consequence, the owners of those kiosks which are served later by other tours will then ask for earlier delivery times, claiming that their kiosk had always opened earlier and that the opening time stated in the wholesaler's list was wrong.

Therefore, when replanning tours, we have to take into consideration that the delivery deadlines of neighboring customers cannot be treated separately, but that they have to be coordinated. We detected this problem in the old town of a metropolis. The routing plan made sure that every customer was served on time. However, after the introduction of the new plan a lot of kiosks in the old town corrected the time of delivery they had originally requested. Especially those kiosks corrected their delivery times which had been served later according to the former plan, but which were now confronted with the situation that a competitor was served earlier according to the new plan.

- *about a figure defining the capacities required for supplying every customer throughout the year.* This data does not present a major problem. Wholesalers can provide data as to the value of their daily deliveries. Delivery notes are produced automatically and kept for half a year to a year. Often wholesalers can also furnish data on the weight of the consignment. In general, we calculate with € 5 per kg of the consignment. Although this may not apply to individual cases, it is obviously a good basis for obtaining an estimated average figure.
- *about the departure times of the tours throughout the year.* These figures are frequently available. Wholesalers often automatically record the times vehicles depart from the depot. In other cases, the times of departure are collected over a certain period of time in order to prepare the planning process. A problem occurs only if the publishers deliver their newspapers in the morning with a delay. In this case, arrangements have to be negotiated between the publisher and the wholesaler as to the maximum period of time a vehicle will wait for a delayed consignment, or whether this consignment will be delivered separately, and who will have to bear the costs of such an extra delivery.
- *providing a distance matrix of the times required to reach every customer from every other customer's location.* This information should depend on the time of the day the tour is effectuated. Obviously, we can only base plans on estimates of these times. Expensive network systems include a categorization of streets. The user can select an average speed according to the categories of the selected streets (Solot / Cuenot / Proca (1990)). However, the average speed also depends on fluctuations caused by the flow of traffic and the weather. Therefore, these networks can only provide for an estimate of the time sequence. In view of these problems, we suggest that the duration of a tour should be estimated by

using a linear function of the Euclidean distance (Becker / Beckefeld / Dillmann (1995) and (1996), Kolesar / Walker / Haussner (1975)).

- *about the time required for unloading the consignment at the customer's location.* This time also varies from day to day and is unknown in most cases. We proposed that time estimates should be based upon a standardized volume of the consignment.
- *about the duration of every tour.* Obviously, the duration varies from day to day. Time estimates for tours are usually based upon the following formula:

  duration of the tour = average speed * distance + stationary time * number of stops + unloading time per parcels * number of parcels

  This formula is used when road networks form the data base. It can be applied to different average speeds for different categories of streets. For better approximations we choose piece-wise linear functions of the Euclidean distances, using different parameters for different lengths. As far as longer distances are concerned, trunk roads or highways can be taken to avoid inner-city traffic and, thus, achieve a higher speed. Furthermore, when separating the actual tour duration we over-estimated the average driving speed and the unloading time. Underestimating the time necessary for driving but over-estimating the time required for unloading serves as a buffer for the replanning process, using the same parameters. The unloading time remains constant, while the travel distance of the new tours is reduced.
- *about the actual sequence of the tours.* This requires a reconstruction of the status quo before replanning the tour, characterized by
  - the daily travel distance
  - the number of customers served with a delay
  - the time by which deliveries to customers are delayed
  - the number of customers on every tour.

  This reconstruction is necessary because although the majority of wholesalers draw up a list defining in which order delivery is to be effected to the customers assigned to a tour, in most cases this order does not correspond to the actual sequence followed by the driver. Moreover, expecting that deviations may be detected by simply looking for zigzag lines in a tour also proves to be misleading. As most tours including time windows show zigzag lines, this is only possible in a small number of cases.

Although reconstructing such a status quo seems to be an easy task, none of the wholesalers was able to provide us with a complete set of the most fundamental data. In some cases the only description of the delivery sequences available to us was the order in which the vehicles were loaded. Thus, we had to reconstruct the status quo ourselves. However, mistakes may lead to an overestimation of the travel distance and an underestimation of the quality of the existing plans. Such mistakes are due to

- an inappropriate positioning of the customers within the network because of different addresses (delivery address and invoice address, special arrangements for individual customers nobody had thought of when preparing the data base)

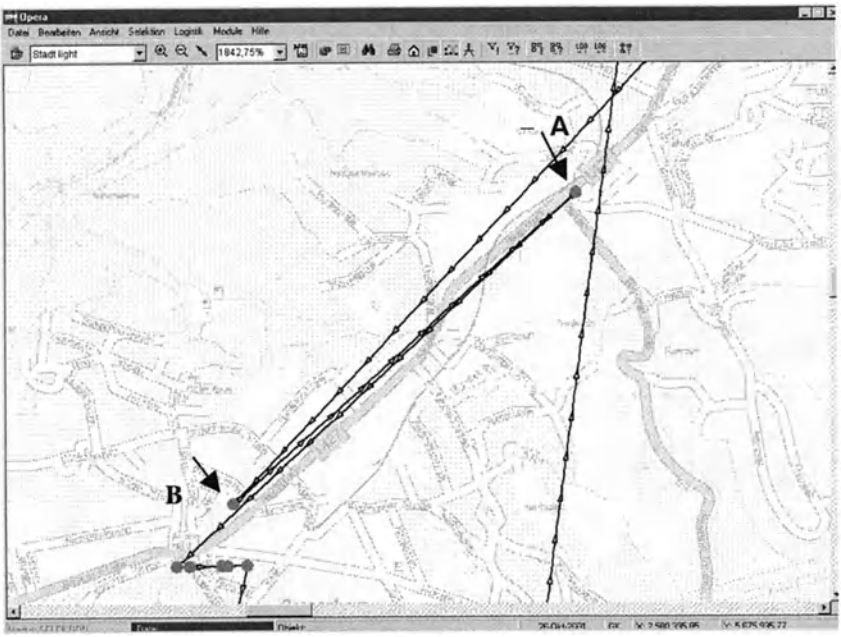


Fig. 2. Zigzag line of a reconstructed tour

- wrong delivery deadlines, especially in case of kiosks and supermarkets with fixed loading ramp intervals (moreover, small hotels can be supplied later, if, for example, the delivery deadline is 5 a.m., but there is only little turnover at that time. When asked whether it was really necessary to deliver so early, these customers said that 6 a.m. was still early enough.)
- unrealistic delivery deadlines for gasoline stations
- a poor knowledge of the details of the actual delivery sequence. (See Fig. 2, which shows the visualization of original sequences stored in the press wholesaler's database. According to the database, customer A is served after customer B. The actual situation, however, is just the opposite: Customer A is served first. In fact, the database shows the packing sequence and not the daily driving sequence.)

Apart from the problem of obtaining and validating the data required for a suitable reconstruction, the persons responsible for logistics at the wholesaler's are sometimes not of much help, either. They often do not correct mistakes. The staff accepted the standard forms without reflecting on the actual status quo.

4 The Planning Philosophy

Computer software for vehicle routing offers several local search methods, and sometimes also metaheuristics as well as a function to change sequences by hand. The methods of local search are based on simple definitions of neighboring tour plans. The complete neighborhood of a solution is checked for possible advantageous changes. Such *general neighborhood* definitions are usually based on the question as to what should be changed and how many tours will be affected. In the relevant literature we can find information about

- *the number of edges changed in one tour.* For example, there is the 2-optimality to avoid crossings (see Croes (1958)); the 3-optimality to move sections within a tour (see Lin / Kernighan (1973)); the Or-optimality concept as a special case of a 3 edges' change (see Or (1976)); a 6 edges' change to change two strings at the best places as special cases of the λ-interchange (see Osman (1993)); and special 4 edges' changes like the 4^*-optimality (see Renaud / Boctor / Laporte (1996a)) - the last one restricts the number of 3 edges' changes and 4 edges' changes by regarding two chains (v_1, v_2, v_3) and $(w_0, w_1,\ldots,w_t, w_{t+1})$, where $u \geq t \geq 1$ for a fixed u. The number of possible 4 edges' changes is reduced from $O(n^4)$ to $O(n^2)$ by fixing the shorter edge of (v_2, w_1) and (v_2, w_t) first, which is feasible if it is shorter than the longest one of the edges (v_1, v_2), (v_2, v_3), (w_0, w_1), and (w_t, w_{t+1}). Let (v_2, w_1) fulfil this condition. Then all remaining 8 possibilities to replace (v_1, v_2), (v_2, v_3), (w_0, w_1), and (w_t, w_{t+1}) by (v_2, w_1) and 3 other edges are examined.
- *how to change several nodes*; Renaud / Boctor / Laporte (1996b) give a 3–tour exchange as an example: Choose node v_1 of tour 1, insert it between v_2 and v_3 in tour 2, select node v_4 of tour 2 where v_4 ? v_2 and v_4 ? v_3 , insert v_4 between v_5 and v_6 of tour 3. This procedure can be interpreted as a special case of two 3 edges' changes or a λ-interchange.
- *a combination of a node-change and a local rearrangement of the schedule* (see Gendreau / Hertz / Laporte (1992) with GENI and US; in GENI, they delete 3 or 4 edges and insert one node and 4 resp. 5 arcs in a prescribed form. In US, they delete 4 or 5 edges and insert 3 resp. 4 edges by excluding one node from the tour. To keep the number of possibilities low, they restrict the procedure to p-neighborhoods. For any node v, its p-neighborhood is the set of the p nodes on the tour closest to v. In case of GENI, it is the p-neighborhood of the node behind which the new node is inserted; in case of US, it is the p-neighborhood of the node to be excluded.)

These and other methods may be combined with metaheuristics to overcome weak local optima. Such methods are

- simulated annealing (see Burkard / Rendl (1984), Cherny (1985), Rossier / Troyon / Liebling (1986), Osman (1993), van Breedam (1995), Marin / Salmerón (1996), Homberger / Gehring (1998))
- tabu search (Glover (1989) and (1998), Glover / Laguna (1995), Glover / Taillard / de Werra (1993), Pesch / Voss (1995), Osman (1993))

- record-to-record (Golden (1993), Dueck (1993))
- Great Deluge Algorithm (Dueck (1993))
- threshold accepting (Dueck / Scheuer (1990))

We feel that such methods are too unspecific to solve typical press distribution problems with many stops, i. e. nodes, as discussed here. First of all, introducing a neighborhood that adequately reflects the problem would lead to an extremely high number of possible candidates. Due to time windows there are only a few admissible solutions in spite of a very high number of candidates in the neighborhood. The task of proposing complex concepts of neighborhoods and restricting the number of possible candidates by introducing adequate rules could not be solved in a satisfactory way. Instead, we implemented multi-objective-oriented automated sequences such as

- the breaking up of complete tours by relaxing the departure times and capacities of other tours; in most cases there is only a small number of tours into which the customers of the resolved tour can be adequately inserted.
- Then, it is necessary to re-sort the customers to equalize the capacities of the tours. This is done step by step. The number of tours to which customers are added is increased, and the number of tours from which customers are removed is reduced. The first tour to which customers are added is the tour with the highest degree of free capacities. In every step we increase the set of tours to which customers are added by the tour with the highest degree of free capacities, which is currently not in the set. If necessary, we repeat the procedure up to three times.
- After that, we re-sort every tour by placing customers opening late at the end in order to be able to effect earlier deliveries to customers opening early.
- Then, we re-sort those customers who are supplied with a delay into other tours.
- Finally, we try to homogenize the numbers of customers supplied by every tour.

While these changes help to achieve a number of objectives, they may also affect the achievement of other goals to a certain extent. This may be due to an acceptance of:

- detours of, for example, 1 km per step (depending on the problem)
- an additional delay of 5 minutes
- 10% overloading
- a maximum number of customers on a tour.

Intensifying those restrictions, which are easier to fulfil, will prevent solutions from being repeated. After some steps involving a guided acceptance of the fact that there may be negative effects on some objectives, there will be a phase of re-optimization, keeping the current status of restrictions. The complete procedure may be repeated several times. The main potential of success of this approach results from the possibility that geographically suitable changes are accepted which could not have been effectuated earlier because there had been other constraints.

Although we know from previous experience that such procedures improve schedules, they do not reach the potentials achieved by dialog. The usual definitions of neighborhoods are based on a change of the maximum number of customers within the same tour or on an exchange between tours. These definitions are space-oriented, i.e. they are motivated by a reduction of distances. But time windows restrict the number of possible changes, which are advantageous from a geographical point of view. Many wholesalers who had tried to optimize their plans with professional systems especially designed for solving transportation problems in case of tours that are subject to daily changes were not able to detect any optimization potentials, because the neighborhood the software was based upon was too simple. These wholesalers were impressed even by small improvements and thought that they had obtained an optimum plan.

To achieve a higher optimization potential, we take advantage of the fact that routing problems can be visualized quite well. We experienced that the human eye is able to recognize those potentials easily where using a mathematical approach for defining an adequate neighborhood concept would be difficult, time-consuming and therefore too expensive.

Neighborhood concepts like

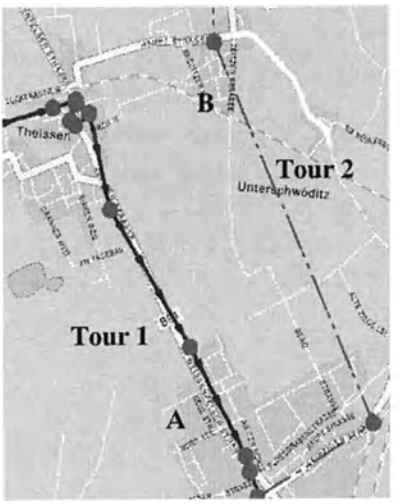
> "re-sort the predecessors of a time-critical customer of tour A as far as possible into other tours by accepting a limited detour and fill the gap by inserting another customer who is producing serious detours in the current tour B and who is located on the way to the time-critical customer in tour A"

or

> "sort part of a cluster of tour A, which can be served by tour B on time and without any major detours, into tour B, so that you can (for capacity reasons) sort into tour A a cluster of customers of tour B located far away from the other customers of tour B and therefore producing considerable detours"

are two examples of improvements, which can be easily detected by visualization. Another example is shown in Fig. 3 and Fig. 4. In Fig. 3 customer A is served too late because this tour does not reach the region in question on time. We sorted the customer from tour 1 into tour 2, which reaches the same region earlier. Customer B could be sorted into tour 1 without causing any time-related problems and any violation of capacity restrictions. This exchange does not cause any major detours in both sequences, so that time buffers are only used up on a limited scale. The customers are served on time.

Such neighborhoods can be defined in many ways. A successful approach depends on the underlying problem. Moreover, defining mathematical functions for such neighborhoods is obviously complicated. We observed that practicians realize this, as well, so that it is not surprising that professional programs are not based on such functions. Instead, we trust our eyes to find out how to choose a particular neighborhood in a given situation.



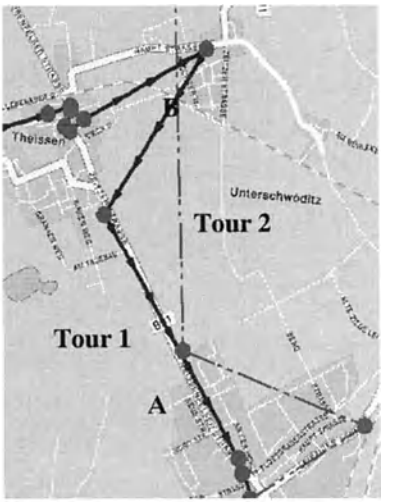


Fig. 3 and **Fig. 4.** Changing sequences

Consequently, improvements based on such neighborhood concepts are best achieved by changing sequences by hand. However, the solution procedure is thus changed fundamentally.

Complex improvement potentials are detected visually and checked by carefully analyzing schedules. Calculations are made in dialog. The decisive and complicated changes are effected by hand. The implemented general neighborhood concepts show the new optimization potentials achieved by manual rearrangements. In fact, such a planning is a carefully directed trial-and-error approach. An analysis of schedules by an experienced person will show the improvement potential that may be achieved by effecting several manual rearrangements. If expectations are not met, return to the original situation, or complete the rearrangement with less success than expected.

Classical heuristics were based on the idea of defining simple and efficient general neighborhood concepts and of examining all possible candidates for 2-changes, 3-changes and so on. We think that our procedure is a consistent further development of a common approach of neighborhoods: We created more complex exchange rules by reducing the number of neighbors and excluding a lot of candidates by a further rule which is illustrated by the work of Renaud / Boctor / Laporte (1996a) and Gendron / Hertz / Laporte (1992). They restrict the number of candidates by using parameters when defining their neighborhood concept. Instead, we use our eyes to detect the candidates who are of interest to us. Thus, we reduce the number of cases to be examined and increase the complexity of the examination in a natural way. There would be no point in introducing these procedures on a general basis. In contrast to the general definitions of neighborhoods, this examination is case-specific. The efficiency of this procedure results from the fact that each situation is handled individually, modifying general ideas in such a way that they can be specifically applied to the individual case.

The procedure proposed by us is recommended in situations where a reduction of the travel distance will reduce costs, due to the fact that the number of routes has been reduced. For example, a reduction of the daily travel distance by 10 km will save 3,000 km a year. Table 2 provides data on a consulting project. The plan developed in the scope of this project leads to an overall cost saving of over € 100,000 per year. It reduces the distance to be covered by about 500 km. Moreover, all customers are supplied on time.

Table 2: Improvements achieved in the course of a planning project

	Before planning[1]	After planning
Number of customers	1,728	1,699[2]
Number of tours	40	37
Total air distance	4,429 km	3,536 km
Total number of customers supplied with a delay	132	0
Total delays in minutes	1,853	0

[1] Including data errors

[2] 29 customers closed down their business during planning

The success of our approach shows that marketing slogans like "optimum efficiency within seconds" are misleading. Since our approach is a time-consuming one, there is a natural end to the replanning process. The payback period of the planning project should be fixed. Then each planning day for which payback is higher than costs is a useful day.

The planning procedure proposed by us offers other advantages, as well. First of all, the practicability of the plan can be controlled at an early stage. Although we schedule on the basis of Euclidean distances, there will be no problem of being misled (lack of bridges, no direct connections), because these barriers are detected by visualizing every tour. This is an important part of the dialog procedure. In addition, we implemented the possibility of introducing artificial nodes between two customers, replacing the straight line by a polygon (see Fig. 5 and Fig. 6).

The user can reproduce those parts of the traffic network judged as important for a better evaluation of the plan without buying a complete network. This is an important aspect of our concept: Unnecessary technical and expensive solutions are replaced by case-specific manual work, physical effort and stamina, giving the user a feeling for the situation immediately.

Secondly, this approach offers a practice-oriented solution. To make sure that the solution reflects the real situation, experienced and trustworthy drivers validate each tour in practice. They rearrange the schedule with regard to one-way streets and other local problems unknown to us. Moreover, the real length in km is determined, which can be checked with the help of standard network-based programs used for calculating the shortest distance between two nodes, or by recalculating the tour distance on the basis of traffic networks. Our tool, which allows for an integration of artificial tour stops, offers another possibility of calculating the length of the tour adequately.

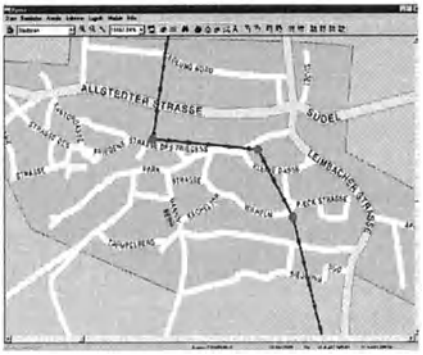


Fig 5. Air distances

Fig. 6. Air distances approximated to the street network

Furthermore, the fact that experienced and trustworthy drivers verify the tours is a useful instrument for achieving an acceptance of the new plans. As drivers participated in the decision-making process, they were able to have an influence on the schedules. They could apply their know-how, especially their good knowledge of the actual traffic situation.

5 Successful Projects and Failures: A Résumé

The approach described in this paper is not suited for problems, which are subject to daily changes, as it remains a time-consuming one. These problems are usually not very complex, because the input data is far clearer and there is no high degree of complexity resulting from the number of stops.

Whether it is recommended to consider taking scientific advice to solve a vehicle routing problem apparently depends on the customer-related degree of complexity (growing number of stops, growing number of constraints, growing danger of neglecting constraints, growing advantage of reducing distances) and the stability of deliveries. The better these requirements are fulfilled, the more likely is a successful introduction of the results of planning in practice.

Reasons for failures:

1. In one case, the delivery deadlines were not correct. We reduced the number of tours by about 20%. The potential of reducing costs while maintaining a good service was enormous because the existing plan was very convenient. But 20% proved to be wrong. This would not have presented a serious problem, as raising the number of tours by 2 tours would have been sufficient to correct the plan. The first day was chaotic; on the second day the company returned to the former plan, without giving it a second try. Once a plan has failed, there will be no second chance any time soon.
2. In another case, the wholesaler saw no possibility of considerably reducing payments to his drivers. Although every tour was overloaded, he refused to accept increasing costs. This was a serious restriction, as we could only shift the

problem from one tour to the next. The problem was over-restricted, and the best thing would have been to refuse to accept the task. There was no way that a new plan would be accepted, as it would have only meant to pass the problem on from one person to the next.

3. A third case showed that increasing the number of tours while keeping the overall distance had to be compensated for by higher total costs in order to maintain a good working climate. The wholesaler tried to keep the costs at their current level, because he was able to reduce the entire daily work, although there were more tours. When the new plan was introduced, costs did not increase. Unfortunately, however, the drivers grew suspicious, so that later on changes were only accepted after lengthy discussions.
4. Another problem came up because the person responsible for logistics had changed. The successor was faced with the problem that if the project were successful, this success would be attributed to his predecessor. He pointed out that he was afraid of implementing the new plan. He did not say why. He refused to control the plan, as in his opinion the tours were not "compact" because they did not follow his concept of regionality. The wholesaler supported this point of view. What we learned from this project was that a good plan does not automatically mean a successful outcome of the project. If the person responsible for such a project changes, it can easily happen that there is no interest in its success any more on part of the successor. Routing is existential for wholesalers, and routing projects cannot be implemented successfully if they collide with the interests of persons in decisive positions. We learned that a mentor exercising a strong influence is crucial for the success of a project.
5. In another case, the printing sequence of a newspaper had to be changed because there had been a fire in the printing company. As a result, the tours started from the depot later. With the original plan, the difficulties would have been much worse, but the delays occurred immediately after the new plan had been introduced, so that retailers attributed the delays to the new plan.

On the other hand, 30 projects were successful, with cost reductions of 5% to 20%. To evaluate our success, it is important to understand how cost reductions could be reached. We always reduced the length and the number of tours. But that was no guarantee for a significant cost reduction. Moreover, replanning the delivery completely offers the chance of introducing a new cost calculation scheme to overcome historically grown payment arrangements. There had often been individual events in the past, which had led to the fact that payments were much higher than necessary. There had been no correction of these unnecessary payments before, as it had not been possible to detect them. Therefore, detecting such costs strongly contributes to the efficiency of our approach. Although there is no direct relation to the quality of the new logistics solution, it is important to define the underlying problem precisely. Against this background, re-defining the cost calculation scheme of a tour was an important task with most of our logistics projects.

Very high returns could be achieved with regard to delivery plans, which had been established within a short period of time, as it was the case in eastern Germany after the German reunification. It seems to us that in western Germany,

where routing plans had been established a long time ago, our main success potential was based on working in dialog and on achieving a fast increase in the awareness of the problem and the knowledge of the region in question in the course of the project. These were the main reasons why we were able to build up confidence among drivers, logisticians and wholesalers.

When changing sequences fundamentally and completely re-organizing daily deliveries, we were not only faced with the problem of mathematical planning, but also of acceptance. Drivers have to be convinced, and the staff managing logistics has to opt in favor of the implementation of a new plan. These people must be given the opportunity to change the plans. It is our task to propose solutions, not to make decisions. Software companies offer tools to find *the* optimum solution. These companies often try to convince people that, by using their software, an optimum solution can be reached with little effort. This, however, seems to be a fundamental error. Our experience has shown that plans that have been gradually developed by hand cannot be matched by the simple heuristics on which software applications and their automated processes are based. Yet, working in dialog is often possible. But an experienced user should do this. Seizing the opportunity of changing an automated solution by hand will lead to a gradual change of the planning process from an automated one to a process that compares with a solution based on scientific advice.

Yet, it was not until the end of the project that we were able to define all the aspects of the individual press distribution problems we intended so solve. Such an insight had not been possible when we started to advise wholesalers.

Acknowledgement

Thanks to Paul Stähly and Andreas Klose for an excellent conference in Sankt Gallen. Thanks also to Simon Goertz and Thomas Bieding for intensive discussions and a lot of valuable comments.

References

Becker, B. / Beckefeld, V. / Dillmann, R. (1994): Tourenplanung im Pressegrosso. In: U. Dehrigs / A. Bachem / A. Drexl, eds., *Operations Research Proceedings*, pp. 547-522. Springer-Verlag, Berlin.

Berens, W. / Körling, F.-J. (1985): Estimating road distances by mathematical functions. *European Journal of Operations Research*, 21:54-56.

Brimberg, J. / Love, R.F. (1991): Estimating travel distances by the weighted I_p norm. *Naval Research Logistics Quarterly*, 38:241-259.

Burkard, R.E. / Rendl, F. (1984): A thermodynamically motivated solution procedure for combinatorial optimization problems. *European Journal of Operations Research*, 17:169-174.

Cerny, V. (1985): Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, 45:41-51.

Christofides, N. / Eilon, S. (1969): Expected distances in distribution problems. *Operational Research Quarterly*, 23:437-443.

Croes, G.A. (1958): A method for solving travelling-salesman problems. *Operations Research*, 6:791-812.

Dillmann, R. / Becker, B. / Beckefeld, V. (1996): Practical aspects of route planning for magazines and newspaper wholesalers. *European Journal of Operations Research*, 90:1-12.

Dueck, G. (1993): New optimization heuristics, the Great Deluge Algorithm and the Record-to-Record Travel. *Journal of Computational Physics*, 104:86-92.

Dueck, G. / Scheuer, T. (1990): Threshold Accepting: A General Purpose Optimization Algorithm Superior to Simulated Annealing. *Journal of Computational Physics*, 90:161-175.

Gendreau, M. / Hertz, A. / Laporte, G. (1992): New insertion and postoptimization procedures for the traveling salesman problem. *Operations Research*, 40:1086-1094.

Glover, F. (1989): Tabu Search. Part I. *ORSA Journal on Computing*, 1:190-206.

Glover, F. (1998): Tabu Search – Wellsprings and Challenges. *European Journal of Operations Research*, 106:221-225.

Glover, F. / Laguna, M. (1995): Tabu Search. In: C. R. Reeves, ed., *Modern heuristic techniques for combinatorial problems*, pp. 70-150, McGraw-Hill, London.

Glover, F / Taillard, E. / de Werra, D. (1993): A user's guide to tabu search. *Annals of Operations Research*, 41:3-28.

Golden, B.L. (1993): Vehicle routing problems and variants. *American Journal of Mathematical and Management Science*, 13:245-248.

Holt, J.N. / Watts, A.M. (1988): Vehicle routing and scheduling in the newspaper industry. In: B.L. Golden / A.A. Assad, eds., *Vehicle Routing: Methods and Studies*, pp. 347-358, Elsevier, Amsterdam.

Homberger, J. / Gehring, H. (1999): Ein Simulated-Annealing-Verfahren für das Standardproblem der Tourenplanung mit Zeitfensterrestriktionen. In: P. Kall / H.-J., Lüthi, eds., *Operations Research Proceedings*, pp.483-492, Springer-Verlag, Berlin.

Kolesar, P. / Walker, W. / Hausner, J. (1975): Determining the relation between fire engine travel times and travel distances in New York City. *Operations Research*, 23:614-627.

Lin, S. / Kernighan, B.W. (1973): An effective heuristic algorithm for the traveling-salesman problem. *Operations Research*, 21:498-516.

Love, R.F. / Morris, J.G. (1972): Modelling inter-city road distances by mathematical functions. *Operational Research Quarterly*, 23:61-71.

Love, R.F. / Morris, J.G. (1988): On estimating road distances by mathematical functions. *European Journal of Operations Research*, 36:251-253.

Marin, A. / Salmerón, J. (1996): A simulated annealing approach to the railroad freight transportation design problem. *International Transactions in Operational Research*, 3:139-149.

Or, I. (1976): *Traveling salesman-type combinatorial optimization problems and their relation to the logistics of regional blood banking*, Ph. D. Thesis, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston.

Osman, I.H. (1993): Metastrategy simulated annealing and tabu search algorithms for the vehicle routing problem. *Annals of Operations Research*, 41:421-451.

Pesch, E. / Voss, S. (1995): Strategies with memories: local search in an application oriented environment, Applied local search – a prologue. *OR-Spektrum*, 17:55-66.

Renaud, J. / Boctor, F.F. / Laporte, G. (1996a): A Fast Composite Heuristic for the Symmetric Traveling Salesman Problem. *INFORMS Journal of Computing*, 8:134-143.

Renaud, J. / Boctor, F.F. / Laporte, G. (1996b): A tabu search heuristic for the multi-depot vehicle routing problem. *Computers & Operations Research*, 23:229-235.

Rossier, Y. / Troyon, M. / Liebling, Th.M. (1986): Probabilistic exchange algorithms and euclidean traveling salesman problems. *OR-Spektrum*, 8:151-164.

Solot, P. / Cuenot, E. / Proca, A. (1990): The distribution of photographic material in Switzerland, *INFOR*, 29:213-224.

Stokx, C.F.M. / Tilanus, C.B. (1991): Deriving route length from radial distances. *European Journal of Operational Research*, 50:22-26.

Van Breedam, A. (1994): *An analysis of the behavior of heuristics for the vehicle routing problem for a selection of problems with vehicle-related, customer-related, and time-related constraints*, Ph.D. Thesis, Faculty of Applied Economics, University of Antwerp.

Van Breedam, A. (1995): Improvement heuristics for the vehicle routing problem based on simulated annealing. *European Journal of Operations Research*, 86:480-490.

Ward, J.E. / Wendell, R.E. (1985): Using block norms for location modeling. *Operations Research*, 33:1074-1090.

Chapter 6

Warehousing

Design of a 2-Stations Automated Guided Vehicle System

Philippe Chevalier[1], Yves Pochet[2], and Laurence Talbot[1]

[1] Institut d'Administration et de Gestion, UCL, 1348 Louvain-la-Neuve, Belgium
[2] CORE, UCL, 1348 Louvain-la-Neuve, Belgium

Abstract. We present analytical results aimed at providing a methodology for the design of a 2-stations Automated Guided Vehicle System (AGVS). The AGVS consists of a number of vehicles transporting products between 2 stations. In this paper, the design of an AGVS consists of determining the dispatching rules and the number of vehicles needed to guarantee some product mean waiting time. The dispatching rules indicate how to well utilize the vehicles. Those rules are provided by using Reorder Point Inventory Policy in order to achieve a fill rate. The fill rate is computed in order to respect the maximum mean waiting time. We use Markov Chain theory to estimate the minimum number of vehicles needed to ensure that the dispatching rules work correctly. Finally, we carry out some simulations to validate our model.

1 Introduction

An Automated Guided Vehicle (AGV) is a driverless vehicle which can accomplish material handling tasks (i.e. load, transport and unload). An Automated Guided Vehicle System (AGVS) consists of a number of vehicles operating in a facility, usually controlled by a computer. The computer takes the dispatching and the routing decisions. AGV technology is a key factor in reducing material handling operating costs and increasing the reliability of material handling systems. However, the purchasing and installation costs are significant, hence the design is an important decision that should be made carefully.

In this paper, we analyze the design of a 2-stations system depicted in Figure 1. This research is originated from a real case, which has to remain confidential. Imagine, for instance a transport system between a workstation and a warehouse. Products are processed at the workstation according to independent processes. After operations at the workstation, the products are transported by vehicles (AGVs) towards the warehouse. The warehouse sends also products loaded on vehicles towards the workstation, when they are required at the workstation. The vehicles are unloaded when they arrive at their destination and the empty vehicles join a storage area.

We seek a model to estimate the minimal number of vehicles needed to guarantee some target service level expressed in terms of mean waiting time. To achieve this we also need to design dispatching rules in order to utilize the vehicles in the best way. Our solution procedure starts by computing the

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

necessary fill rate in order to respect the maximum mean waiting times. The dispatching rules and the minimum number of vehicles are determined to guarantee the fill rate.

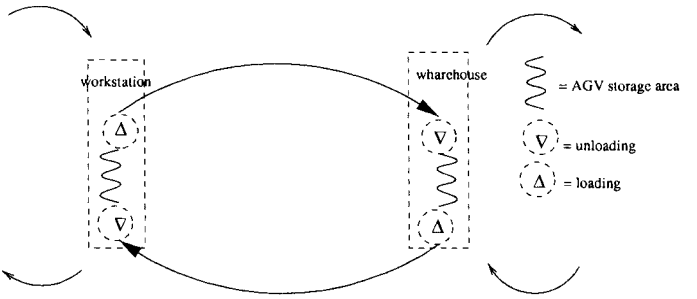


Fig. 1. Environnement

The content of the paper is as follows. Section 2 presents the AGV environment and the motivations of the analysis. Section 3 presents a literature review for similar AGVS design problems. In Section 4, we present our solution procedure. In section 5, we evaluate the quality and the relevance of our approach. We conclude in section 6.

2 AGV Environment and Research Motivations

Consider a 2-stations system depicted in Figure 2. Products arrive at each loading station to be transported to the other station. For each station i the arrival process is characterised by a mean interarrival time of $\frac{1}{\lambda_i}$ and a variance of σ_i^2. A product waits in queue until a vehicle is available to transport it towards the other station. When arriving at the other station, the product is unloaded and leaves the system (and thus the environment analyzed). The emptied vehicle joins the storage area and becomes available for a transportation request at the other station.

We introduce the following assumptions:

- The interarrival times form a renewal process (interarrival times are IID random variables).
- The arrival processes of the two stations are independent.
- The queue of products at the two stations have infinite capacity.
- Only one product can be transported by a vehicle at a time.
- The products are dispatched according to the First-Come-First-Served rule.

- The loading time and the unloading time are zero.
- The internal travel time within the station is neglected.
- We denote by Δ the minimum time separating the sending of two vehicles at a station (technology limit).
- The travel time T between the two stations is known with certainty and is independent of the load being conveyed by the vehicle.
- The vehicle storage area at each station is unlimited.
- We will only study the system in its steady state.

The products are supposed to be transported as soon as possible. The objective is to minimize the time separating the entry of the product in the waiting queue at one station and the unloading of this product from a vehicle at the other station. Since the travel time T between the two stations is constant, it is equivalent to minimize the product waiting time before loading. Hence, the level of performance is measured by the mean product waiting time at station i, denoted by Wq_i.

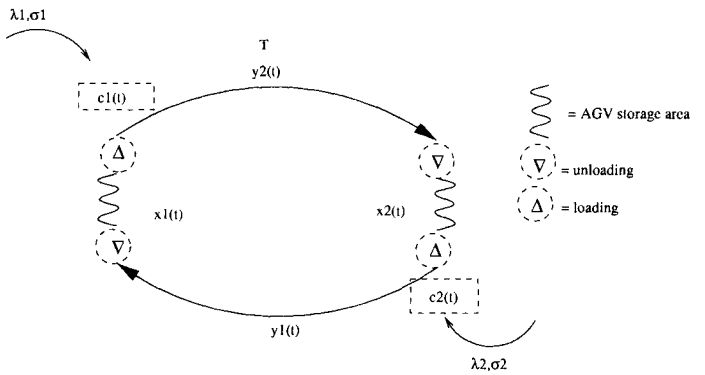


Fig. 2. Environnement

We introduce the following notation:

$x_i(t)$ = The number of vehicles in the storage area of station i at time t.
$y_i(t)$ = The number of vehicles (loaded or empty) traveling towards station i at time t.
$c_i(t)$ = The number of products waiting at station i at time t.
$N_i(t)$ = The cumulated number of products that arrived from time 0 to time t at station i.
X = The total number of vehicles in the system = $x_1(t)+y_1(t)+x_2(t)+y_2(t)$, for all t.
$i-$ = Station preceding station i.

The purchasing and the installation costs of vehicles are often significant. So we seek a model to find the minimal number of vehicles X needed to guarantee some predefined minimal service level expressed in terms of mean waiting time for transportation requests.

To be able to determine this minimal number of vehicles, we first need to determine how to best utilize the vehicles. Therefore we must determine dispatching rules. Dispatching rules are used to manage the stock of vehicles at the two stations. In order to prevent a shortage of vehicles at a station to cover transportation requests, the dispatching rules determine whether to send empty vehicles from one station towards another. In this way, available vehicles are divided among stations to provide the best service level.

2.1 Research Motivations

In this section, we explain how the number of vehicles and the dispatching rules have an influence on the mean waiting time.

First of all, the number of vehicles will never need to exceed $\frac{2T}{\Delta}$ because $2T$ is the total transportation time for a round trip and Δ is the minimal time interval between two departures of vehicles (technology limit). Therefore, a lower bound on the mean waiting time is observed when one vehicle leaves each station every Δ units of time. This lower bound is the mean waiting time of a $G/D/1$ system with service time Δ, denoted by $\mathrm{WD}(\Delta)$. If the number of vehicles is $\frac{2T}{\Delta}$ the lower bound $\mathrm{WD}(\Delta)$ can easily be reached by sending the vehicles at regular intervals, one every Δ units of time. But, the objective is to minimize the number of vehicles.

Similarly, when the number of vehicles is X, it is possible to send one vehicle from each station every $\frac{2T}{X}$ units of time. Therefore, we observe that the mean waiting time of a $G/D/1$ system with service time $\frac{2T}{X}$, denoted by $WD(\frac{2T}{X})$ is a level of performance easily reached. $WD(\frac{2T}{X})$ is an upper bound on the optimal mean time for a given number X of vehicles.

Moreover, we observe that the role and the impact of the dispatching rules and the number of vehicles depends on the distribution of the transportation demand and its evolution over time. In particular, let us assume that ε is the duration of a peak of transportation demand. Figure 3 represents possible demand evolution over time where the demand evolution is measured by the instantaneous demand rate. The first case is characterized by a lot of randomness. This is the case, for instance, when the interarrival time process follows an exponential distribution. In the second case, the distribution is regular. This is the case when the interarrival times are relatively constant and correlated (which is contrary to our IID assumption).

If the demand peak does not last a long time (small ε) it is possible to constitute a stock of vehicles before the beginning of the peak in order to cover the demand peak. A minimal number of vehicles is required in order to have enough vehicles in stock at the stations to cover these peaks. The

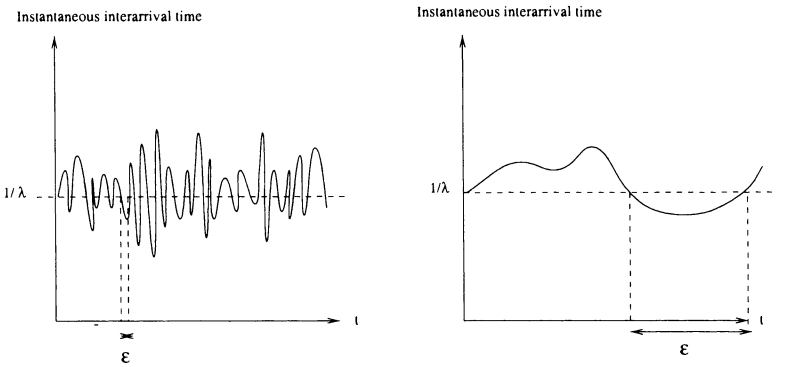


Fig. 3. Demand evolution over time

dispatching rules have the purpose of constituting this stock, and the quality of the dispatching rules is measured in terms of mean waiting time in this case.

If the demand is regular (large ε), the dispatching rules are not useful because there are not enough vehicles to stock and cover the peak. Therefore, the upper bound on the minimal mean waiting time will be reached by sending the vehicles regularly, every $\frac{2T}{X}$ units of times. It corresponds to the mean waiting time of a $G/D/1$ system with service time $\frac{2T}{X}$, denoted by $WD(\frac{2T}{X})$.

Figure 4 represents for a given number of vehicles, the evolution of the mean waiting time as a function of the length of the peak. A demand with a small ε corresponds to a demand with a lot of randomness such a Poisson process, a demand with a large ε corresponds to a demand with relatively constant and correlated interarrival times. The mean waiting time approaches $WD(\Delta)$ when the peak demand is short and approaches $WD(\frac{2T}{X})$ when the demand is regular or the peak is long. Between these two extreme cases, the mean waiting time depends on the quality of the dispatching rules. Curve 2 represents the performance of a better dispatching rule than that of Curve 1.

In conclusion, if the demand is regular, the number of vehicles determine the performance of the system. If there is a short peak of demand, the dispatching rules and the number of vehicles are both important. We want to determine the number of vehicles and the dispatching rules in order to achieve a given performance for given demand distribution. Since this model is supposed to operate in a highly stochastic environment, we anticipate that the dispatching rules will play an important role.

3 Literature Review

The purchasing and installation costs of AGVS are significant, and hence the design is an important decision that should be made carefully. For a given

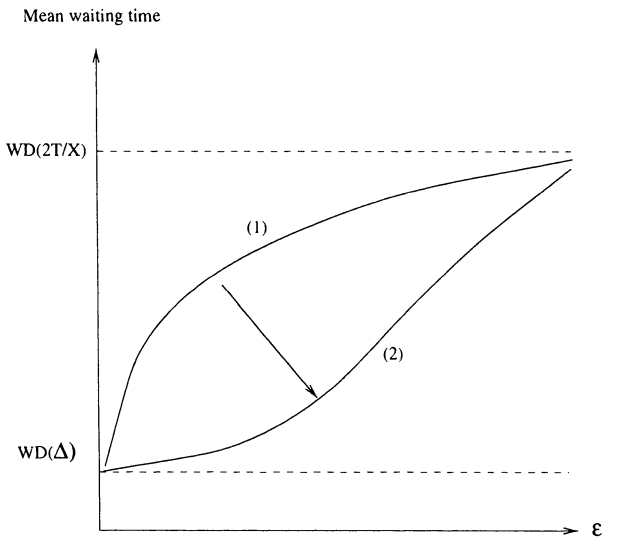


Fig. 4. Evolution of the mean waiting time as a function of the peak length

network layout, AGVS design is primarily concerned with the determination of the number of vehicles needed. The required number of vehicles is affected by several factors: number of transportations, travel time, network layout,...

Newton (1985) estimates the number of vehicles needed in an AGVS by means of a simulation experiment. Wysk, Egbelu, Zhou and Ghosh (1987) and Egbelu (1987) use simulation to test the quality of solutions from analytical deterministic models based on the total transportation distance. van der Meer and de Koster (1997, 1998) used a simulation model to compare decentral control (the vehicle drives in assigned loops from workstation to workstation) with central control (a central controller is used for dispatching of loads or vehicles). They also classified different dispatching rules for uni-load vehicles. It was concluded that central control outperforms decentralized control. van der Meer and de Koster (1999) look at the robustness of the classification when multi-load vehicles are used, and also see how the classification holds up when loads are released in batches at the receiving station. In those papers, the vehicle is sent to a workstation when it is free. They use a "push" approach. We use a "pull" approach; when a workstation needs a vehicle, it calls one. Moreover, we try to provide an analytical model and use simulation only to validate our model.

According to Askin and Standbridge (1993), considerable theory exists aiding the determination of the number of vehicles needed to support the material handling requirements. Although the actual problem is stochastic,

most of them use a simpler deterministic model to estimate requirements. These models estimate the number of vehicles needed as the ratio between the total vehicle utilization time and the effective time a vehicle is available. The total utilization time can be divided into five components: loaded travel time, load time, unload time, unloaded travel time and waiting and blocked time.

Rajotia, Shanker and Batras (1998) describe the analytical model usually employed for the loaded travel time estimation. The loaded travel time required can easily be determined from the information available on material flow intensities and travel time matrix among various pairs of pick up/delivery (P/D) stations. Likewise, specifications on load and unload times can be multiplied by the number of loads to find total load and unload times.

Maxwell and Muckstadt (1982) did pioneering work in analytical modeling of operation features of an AGVS. They estimate for instance the empty vehicle travel time by computing the net flow at each P/D station as the difference between the total number of unit loads delivered there and the total number of unit loads picked up from there. It represents the number of empty trips into or out of that station. A standard transportation problem was formulated which assigned empty trips between various stations minimizing the total empty vehicle travel time. Authors such as Srinivasan, Bozer and Cho (1994), Mahadevan and Narendran (1990), Lin (1990), Malmborg (1990) proposed similar approaches to estimate the loaded and unloaded travel times and the number of vehicles. Askin and Standbridge (1993) discussed several research studies for the purpose of empty travel time estimation and proposed a new model. The model begins with the objective of minimizing empty vehicle trips. the frequency of such trips from/to a station is constrained by the total number of loads delivered at/picked up from that station. Because of these additional constraints, the number of empty trips is greater than the net flow as considered by Maxwell and Muckstadt (1982). They used simulation to validate the approach and the results indicate that the model underestimates the minimum AGV requirements but provides results close to the simulation results. Mahadevan and Narendran (1991) take into consideration the limited local buffer. Mahadevan and Narendran (1994) demonstrate the use of a two stage approach for determining the number of vehicles, the layout of tracks, the dispatching rules for the vehicles and the provision of control zones and buffers. The required number of vehicles is estimated using the model presented in Mahadevan and Narendran (1990) in the first stage. In the second stage, the effects of AGV failures and AGV dispatching rules on the system performance are observed through simulation studies.

The last component of vehicle utilization is waiting and blocking time. Waiting time refers to the time an AGV spends while waiting empty at a delivery station for assignment of the next load transportation task. Blocking time is the time an AGV remains in a blocked state because of traffic congestion while it is traveling either loaded or empty. According to Askin and Standbridge (1998), it is not possible to compute such waiting and block-

ing times in advance because they are dependent upon the AGV fleet size, a parameter which has to be determined first. They depend also upon the dispatching and routing strategies, guidepath layout, and vehicles clearance procedures at crossings. Koff (1987) presented empirical approximation of vehicle idleness (empty travel time and idle waiting time) and blocking time factors. The range of values suggested is 10 to 15% of total loaded vehicle travel time for each of the two.

In all these approaches, the number of vehicles is determined using deterministic approaches based on the mean transportation load. The performance is usually observed by simulation. Our approach is to impose a priori a system performance target, and develop analytical models to compute the number of vehicles required to achieve it, taking into account the stochastic aspect of the demand. Simulation is only used to validate our approach.

Closer to our methodology, Tanchoco, Egbelu and Taghaboni (1987) model the AGVS as a closed queuing network in order to determine the minimum number of vehicles required. The effectiveness of this modeling approach is compared to a simulation based method. Since their analysis ignores the time a vehicle spends traveling empty, it generally underestimates the actual number of vehicles required. Mantel and Landeweerd (1995) developed a hierarchical queuing network approach to determine the number of AGVs.

Johnson and Brandeau (1993) propose analytical models for the design of an AGVS. Taking into account the stochastic nature of the demand process, they apply results from queuing theory to estimate system congestion. They investigate an AGVS that delivers material from a central storage depot to shop floor workcenters. The objective is to determine which workcenters to include in the AGV network and the number of vehicles required to service those workcenters, in order to maximize the benefit of the network, subject to a constraint that the average waiting time at each station does not exceed a predefined limit. The benefit of each station is defined as the net present value of direct labor savings of delivering material via the AGVS minus the cost of constructing a pickup and dropoff station at that workcenter. Their contribution is to solve the problem using an analytical model. The pool of vehicles is modeled as an $M/G/c$ queuing system and the design model is formulated as a linear binary program with a set of nonlinear constraints to model average waiting time at each station. The constraints are expressed by an approximating queuing formula. The approximation attempts to adjust the $M/M/c$ formula to account for service time variations. Thonemann and Brandeau (1997) introduce also an analytical model for the design of a multiple-vehicle AGVS with multiple-load capacity operating under a "go-when-filled" dispatching rule. The AGVs deliver containers of material from a central depot to workcenters throughout the factory floor. The demand of the workcenters and the time until delivery are stochastic. They develop a nonlinear binary programming model to determine the optimal partition of workcenters into zones, the optimal number of AGVs to purchase, and the optimal subset of workcenters to service by the AGVS, subject to constraints

on maximum allowable mean waiting time for material delivery. They develop an analytical expression for the mean waiting time and present an efficient branch-and-bound algorithm that solves the AGVS design model optimally.

Johnson and Brandeau (1993) and Thonemann and Brandeau (1997) are closer to our methodology because they propose an analytical model taking into account the stochastic nature of the demand process and they apply results from queuing theory. But in our paper, we analyze and suggest dispatching rules in order to "best" utilize the vehicles to achieve a given mean waiting time with a smaller number of vehicles.

Almeida and Kellert (2000) study job shop like flexible manufacturing system (FMSs) with a discrete material handling device and machine transfer blocking. They propose an analytical queuing network model to evaluate the quantitative steady-state performance of such FMSs. The FMS complex devices are structured in order to prevent deadlocks from occurring. So, in this paper, they suppose a light load. In our paper, the system is heavily loaded. Bozer and Kim (1996) determine optimal or near optimal transfer batch sizes in manufacturing systems and develop an analytical relationship, issued from queuing theory, between the material handling capacity and the expected work in process in a manufacturing system. The models developed by Almeida and Kellert (2000) and Bozer and Kim (1996) are not applicable to our problem because the aim is not the same. But the methodology is identical. They present an analytical model based on queuing theory and the results are validated against discrete event simulations.

AGVS design is also concerned with dispatching rules.

According to Taghaboni and Tanchoco(1988), dispatching involves the selection rule, or methodology, that is used for selecting a vehicle for pick up of delivery assignments. The problem addressed here is not exactly the same. More precisely, the dispatching rules are used here to organize or distribute the stock of empty vehicles in order to best cover the transportation demands. Due to the simple structure of the transportation system, the dispatching rules are not used to assign vehicles to specific transportation requests.

4 AGVS Design Model

4.1 Framework

The level of performance is measured by the product mean waiting time at the 2 stations, denoted by Wq_1 and Wq_2. As the purchasing and the installation costs of AGVs are significant, we try to minimize the number of vehicles needed to guarantee the level of performance. So we have to good utilize the vehicles. For that, we determine the dispatching rules which guarantee a certain level of stock at the two stations. The dispatching rules are determined by using a classical Reorder Point Inventory Policy (RPIP) from inventory management theory. But, this policy is based on the concept of fill rate and

not on the concept of mean waiting time. The fill rate τ is the probability to have an available vehicle at the station to satisfy a request, at the moment the request occurs.

So, in the first step of our model, we have to provide an expression for the mean waiting time in terms of the fill rate τ in order to be able to use the RPIP to determine the dispatching rules. Finally, we determine the number of vehicles necessary to be sure that the dispatching rules work correctly.

4.2 Finding the Appropriate Fill Rate

We try to provide an expression for the mean waiting time Wq in terms of the fill rate τ. Each station is considered here independently and can be modeled as a queuing process. The independence hypothesis will be satisfied if there are enough vehicles in the system. The products enter a waiting queue according to some interarrival times distribution with a mean interarrival time of $\frac{1}{\lambda}$ and a variance of σ^2. To estimate the waiting time we use the following approximation proposed by Marchal (1978) for a *G/G/1* system:

$$Wq = \frac{\lambda(\sigma^2 + Var(S))}{2(1-\rho)} \frac{\rho^2 + \lambda^2 Var(S)}{1 + \lambda^2 Var(S)} \tag{1}$$

where S is the service time distribution (time between successive departures of vehicles from a station), Var(S) is the variance of S, $\rho = \lambda * E(S)$ and $E(S)$ is the expected service time.

So, in order to calculate Wq, we need to estimate the expected service time $E(S)$ and the variance $Var(S)$. We assume here that the dispatching rule used will ensure a fill rate τ. Thus when a request arrives, there is a probability τ of having at least one vehicle available and a probability $1 - \tau$ that the product should wait for the arrival of a vehicle.

1. With a probability τ, at least one vehicle is available to transport a product. In this case, the service time is distributed according to a constant distribution of parameter Δ (technological limit) because one product is sent to the other station every Δ units of time.
2. On the other hand, with a probability $1-\tau$, there is no vehicle available at the station when a product arrives. The product has to wait until a vehicle arrives. In the worst case, the second station does not send any product. It sends only empty vehicles when the first station requires one. The first station requires a vehicle each time a product arrives, consequently the time between arrivals of empty vehicles will have the same distribution as between arrivals of a product at that station. Therefore, the distribution of the arrival of empty vehicles at the first station, and the service time distribution, are the same as the arrival distribution.

So, together for the two cases, the expected service time is

$$E(S) = \tau\Delta + (1-\tau) * \frac{1}{\lambda}$$

As $E(S^2) = \tau\Delta^2 + (1-\tau) * (\sigma^2 + \frac{1}{\lambda^2})$, the variance of the service time distribution is

$$Var(S) = E(S^2) - E(S)^2 = \tau(1-\tau)(\Delta - \frac{1}{\lambda})^2 + (1-\tau)\sigma^2$$

Knowing $E(S)$ and $Var(S)$, for a maximum waiting time Wq^*, we determine the appropriate fill rate τ^* as

$$\tau^* = min\{\tau : Wq(\tau) \leq Wq^*\}$$

The expression of Wq has been developed in terms of the fill rate τ. Wq depends on τ, Δ, λ and σ^2, assuming that the fill rate τ can be guaranteed.

Now, we want to find the dispatching rules needed to achieve the fill rate τ.

4.3 Dispatching Rules

Any request for a vehicle at station i has to wait on average Wq_i. When station 1 asks for an empty vehicle, it has to wait on average $T + Wq_2$ units of time for the vehicle to arrive. This means that to ensure a fill rate τ at station 1, we must make sure with probability τ, that at any time there are enough vehicles traveling to station 1 to cover the transportation demand (of station 1) during $T + Wq_2$ units of time.

Recall that $N_i(t)$ is the cumulated number of products that arrived from time 0 to time t at station i. If the current time is t, $(N_1(t+T+Wq_2) - N_1(t))$ is the number of products that will arrive at station 1 during the next $T+Wq_2$ units of time. We want to make sure that the number of vehicles traveling to station 1 is high enough so that the probability of the number of arrivals being greater does not exceed $1-\tau$. We call S_1 the smallest number which satisfies the inequality:

$$P\big((N_1(t+T+Wq_2) - N_1(t)) \geq S_1\big) \leq 1-\tau$$

We want to have at least S_i vehicles available at station i at any time. We have the same expression for the station 2. In general, for the station i, S_i is the smallest number which satisfies the following inequality:

$$P\big((N_i(t+T+Wq_{i-}) - N_i(t)) \geq S_i\big) \leq 1-\tau \qquad (2)$$

In other words the probability of not having enough vehicles (i.e. more demands than S_i) is at most $1-\tau$. The dispatching rules try to anticipate the possible demand by calling empty vehicles if necessary. So, we introduce the concept of net stock of available vehicles for station i at time t, $s_i(t)$, which is

defined as the sum of vehicles at station i plus the vehicles *en route* to station i and minus the products waiting at station i (corresponding to vehicles not available to transport new requests). Formally, $s_i(t) = x_i(t) + y_i(t) - c_i(t)$. The net stock $s_i(t)$ has to be higher than S_i in order to satisfy the fill rate τ. This suggests the following dispatching rule:

When the net stock of vehicles at station i, $s_i(t)$, is less than S_i, the other station sends an empty vehicle to station i.

This dispatching rule is similar to a classical reorder point policy from inventory management theory.

4.4 Number of AGVs

Our dispatching rules suppose there are enough vehicles so that a request for an empty vehicle waits on average Wq_i at station i. We will now try to determine the minimum number of vehicles required to ensure this.

First Observation It is obvious that the number of vehicles must be higher or equal to $S1 + S2$. Indeed $X = x_1(t) + y_1(t) + x_2(t) + y_2(t)$ and the net stock $s_i(t) = x_i(t) + y_i(t) - c_i(t)$ must be at least S_i, for $i = 1, 2$.

We need more than $S1+S2$ vehicles. Indeed, if the system has just $S1+S2$ vehicles, there is a risk to see the vehicles running without interruption only to satisfy the dispatching rules (empty transportation requests). If the station 1 sends a vehicle, its net stock becomes less than $S1$, hence the station 1 ask for a vehicle. Station 2 sends this vehicle and its net stock becomes less than $S2$, and so on. This does not allow to satisfy the fill rate requirement.

Number of Vehicles *en route* To achieve a fill rate τ, we need - with probability τ - at least one vehicle available at each station. So, we want, with probability τ, more vehicles in the system than the vehicles *en route*. If we know the maximum number of vehicles *en route*, we can deduce the number of vehicles required.

The number of vehicles *en route* is the number of vehicles sent by the two stations. A station sends loaded vehicles at rate λ_i (λ_i is given) and empty vehicles at the rate ν_i (according to the dispatching rules). If we know ν_i, we assume that the time between successive vehicles on a link (between two stations) has the same type of distribution as the interarrival time process of demand, with the distribution being scaled to take the empty vehicles into account. That is,

$$f_i^l(x) = f_i^a(x * \frac{\lambda + \nu}{\lambda})$$

where

- $f_i^l()$ is the probability density function of times between vehicles on a link.
- $f_i^a()$ is the probability density function of arrivals.

Let $N_i'(t)$ be the cumulated number of vehicles that arrived at station i according to the distribution $f_i^l()$ from time 0 to time t. We want, with probability τ, more vehicles in the system than the number of vehicles *en route*. In other words, we want with a probability $1-\tau$, more vehicles *en route* than the number of vehicles in the system. The number of vehicles *en route* at time t between station $i-$ and station i is defined by $N_i'(t+T+Wq_{i-})-N_i'(t)$. So we determine S_i' for each link to be the smallest number satisfying:

$$P\big((N_i'(t+T+Wq_{i-})-N_i'(t))\geq S_i'\big)\leq 1-\tau \tag{3}$$

If we adopt the same reasoning for the two stations, we can thus say that, with probability τ, there is maximum $S_1'+S_2'$ vehicles *en route*. In conclusion, knowing the rates of empty vehicles (ν_1 and ν_2), the minimum number of vehicles necessary to cover the total transportation is thus $S_1'+S_2'$. In the following section, we develop an expression for ν_1 and ν_2.

Estimation of ν_i We suppose that a station can send a vehicle as soon as the other station requires one. To estimate the rate at which station i sends empty vehicles to the other station, we have to know when the other station requires an empty vehicle. A station requires an empty vehicle when its net stock goes below the limit S_i. Let us define n as the number of vehicles above the minimum of S_1+S_2, in other words: $n=X-S_1-S_2$. We define a stochastic process $E(t)=s_1(t)-S_1$. $E(t)$ is the number of vehicles above the dispatching rule limit at station 1. $E(t)$ will increase by one unit each time a product is sent from station 2 to station 1. $E(t)$ will decrease by one unit each time station 1 sends a product to station 2. When $E(t)=0$ and station 1 sends a product then the net stock would go below the minimum threshold of the dispatching rule and consequently an empty vehicle should be sent from station 2 to station 1. We have the same situation at the other direction when $E(t)=n$.

We will approximate the stochastic process $E(t)$ by a markov process $M(t)$ where we assume the times between transitions are exponentially distributed with parameter λ_1 and λ_2 respectively. This corresponds to Poisson arrival processes for the products. Our simulation results (an example is given in section 5.1.3) show that this approximation gives satisfactory results. Figure 5 represents this Markov process.

The first station requires an empty vehicle only when the markov process is in state 0 and a transportation request arrives to station 1. Station 2 requires an empty vehicle when the system is in the state n and a transportation request arrives to station 2. So, the expression of ν_1 and ν_2 are:

$$\nu_1 = P_n * \lambda_2 \tag{4}$$

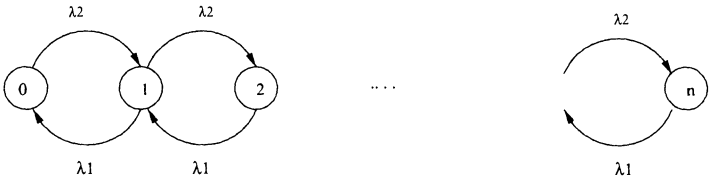


Fig. 5. Markov Process

$$\nu_2 = P_0 * \lambda_1 \tag{5}$$

Knowing λ_1 and λ_2, we have to find P_0 and P_n. If the demands are the same at the two stations ($\lambda_1 = \lambda_2$), all states are equiprobable, and $P_0 = P_1 = P_n = ... = \frac{1}{n+1}$. If the demands are different at the two stations, then $P_n = \rho^n * P_0$ and $P_0 = \frac{1-\rho}{1-\rho^{(n+1)}}$ where $\rho = \frac{\lambda_2}{\lambda_1}$, assuming without loss of generality that $\lambda_2 \leq \lambda_1$.

Conclusion We now have two equations in order to find the number of vehicles X. The first one finds ν_i from n and the second one estimates X from ν_i. We solve this system in an iterative way. We propose the following procedure.

1. For a given mean waiting time Wq, we compute the corresponding fill rate τ (expression 1).
2. For a given fill rate τ, we compute $S1$ and $S2$ (expression 2).
3. We pose n and we compute ν_1 and ν_2 (expression 4 and 5).
4. We compute $S1'$ and $S2'$ and $X = S1' + S2'$ (expression 3).
5. If $X - (S_1 + S_2) > n$, increase n and go back to the step 2.
   If $X - (S_1 + S_2) < n$, decrease n and go back to the step 2.
   If $X - (S_1 + S_2) = n$, n is optimal, go to step 5.
6. $X = S1 + S2 + n$

We did not study the theoretical convergence property of this procedure. But a very small number of iterations were needed in all test cases.

5 Evaluation of the Quality and the Relevance of the Model

In this section, we evaluate the quality and the relevance of our approach. First, we carry out simulations of the system using the Extend Software[1]. Then, we show that a deterministic approach or a pure queuing theory approach can not achieve results of similar quality.

[1] Extend v4.1.3, Imagine That, Inc., 1998, San Jose

5.1 Evaluation of the Quality of the Model

We carried out four series of simulations. The travel time is 100 units of time and the technology limit Δ is one unit of time. The simulations vary in terms of interarrival time processes. The interarrival time distributions are characterized by a lot of randomness (see our IID assumption) and we know the mean and the variance of the distributions. In the first simulation, the interarrival time processes at the two stations are exponential and asymmetric. In the second simulation, the interarrival time processes are exponential and symmetric. In the third simulation, the interarrival processes are non-exponential and asymmetric. In the fourth simulation experiment, we evaluate the robustness of our solution.

We compute, for different fill rates, the mean waiting time (Wq), the dispatching rules (S_i) and the number of vehicles (X) needed. In the simulation, we compare the observed fill rate (τ_{obs}) to the given fill rate (τ) and the observed mean waiting time (Wq_{obs}) to the computed mean waiting time (Wq).

The three series of simulations relate to 10 runs of 10.000 units of time. The tables below present the computed and simulated quantities with their 95% confidence interval.

1. **Asymmetric exponential demands**
   The interarrival time processes at the two stations follow an exponential distribution, with a mean of 2 time units at station 1 and 4 time units at station 2. The simulation results are given in Table 1.

Table 1. Asymmetric exponential processes

objective	design			prediction		observation			
τ	S_1	S_2	X	Wq_1	Wq_2	$\tau_1$obs	$\tau_2$obs	Wq_1obs	Wq_2obs
99%	67	37	134	0.54	0.22	0.99	0.99	0.51±0.04	0.4±0.0256
95%	62	34	124	0.71	0.45	0.97±0.02	0.97±0.06	0.7 ±0.23	0.51±0.05
90%	60	32	120	0.94	0.76	0.94±0.01	0.95±0.01	0.84±0.12	0.6±0.07
85%	58	31	116	1.21	1.1	0.88±0.02	0.92	1.24±0.28	0.83±0.15
80%	57	30	114	1.5	1.5	0.83±0.045	0.9±0.02	1.96±0.93	0.95±0.2

For instance, for a fill rate of 90%, station 2 sends a vehicle to station 1 if the net stock of vehicles at station 1 is less than 60 vehicles. The simulation presents a fill rate of 94% and 95%, better than what we expected. The mean waiting time calculated (0.94 and 0.76) seems to be a good approximation of the mean waiting time simulated (0.84 and 0.6). Since the dispatching rules are mainly used by the station with the highest load, the results are better for that station. The model at that station is closer to reality.

To achieve a fill rate of 100%, the number of vehicles has to be $\frac{2T}{\Delta}$ and the best level of performance $WD(\Delta)$ is reached by sending the vehicles at regular time intervals. In this case, $\frac{2T}{\Delta} = 200$ for a mean waiting time of 0.54. Our model proposes 134 vehicles for a fill rate of 99% and a mean waiting time of 0.94 units of time. This result is interesting knowing that the purchasing and the installation costs of vehicles are often significant.

2. **Symmetric exponential demands**
The interarrival time processes at the two stations follow exponential distribution, with a mean of 2 time units. The corresponding results are presented in Table 2.

Table 2. Symmetric exponential processes

objective	design			prediction		observation			
τ	S_1	S_2	X	Wq_1	Wq_2	$\tau_1$obs	$\tau_2$obs	Wq_1obs	Wq_2obs
99%	67	67	144	0.54	0.54	0.99±0.04	0.99	0.6±0.02	0.61±0.04
95%	62	62	134	0.71	0.71	0.96±0.01	0.96±0.01	0.72±0.01	0.73±0.09
90%	60	60	130	0.94	0.94	0.92±0.02	0.92±0.02	1±0.05	0.96±0.14
85%	58	58	126	1.2	1.2	0.89±0.01	0.89±0.02	1.27±0.29	1.26±0.23
80%	57	57	124	1.5	1.5	0.84±0.02	0.84±0.03	1.36±0.19	1.51±0.3

The observed fill rates are larger than what we expected in our analysis. The mean waiting time calculated seems to be a good approximation of the mean waiting time simulated.

3. **Non-Poisson process**
The interarrival times of transportation requests follow a log-normal distribution, called X_n with a mean time of 2 units of time for station 1 and a variance of 2, and a mean time of 4 units of time and a variance of 8 for station 2. Table 3 presents the results.

Table 3. Non-Poisson processes

objective	design			prediction		observation			
τ	S_1	S_2	X	Wq_1	Wq_2	$\tau_1$obs	$\tau_2$obs	Wq_1obs	Wq_2obs
99%	63	35	126	0.27	0.1	1±0	1±0	0.1±0.01	0.19±0
95%	59	32	118	0.34	0.2	0.98±0	0.98±0	0.19±0.06	0.25±0.02
90%	57	30	114	0.44	0.33	0.94±0.01	0.96±0	0.35±0.1	0.36±0.02
85%	55	29	110	0.56	0.49	0.86±0.01	0.92±0	0.94±0.1	0.52±0
80%	55	28	110	0.7	0.68	0.86±0.02	0.92±0	1.02±0.19	0.52±0.03

The observed fill rates are larger than what we expected for the large fill rates. According to Marchal (1978), the performance of the mean waiting

time approximation deteriorates as the service times or interarrival times deviate further from exponentially. Moreover, to compute the dispatching rules limits S_i, we utilised the Central Limit Theorem. It is known that this approximation will be less precise for the tails of the distribution.

4. **Robustness**
In real cases, the interarrival mean time varies with time $\frac{1}{\lambda(t)}$ around the nominal rate $\frac{1}{\lambda}$. In our approach, we determine dispatching rules and the number of vehicles with a nominal rate, but if the mean varies with time, the level of performance would not be too degraded. We expect that the degradation will be proportional to the difference between $\lambda(t)$ and λ. We have simulated the worst case in which dispatching rules and the number of vehicles are based on exponential interarrival time processes with an interarrival mean time of 2 and 4 units of time, respectively at the two stations, but the real interarrival times have a mean of 4 and 2 units of time, respectively. In other words we expect the heavy flow to be from 1 to 2, but we observe the opposite. We see that the level of performance (Wq) is not too degraded. The simulation results are given in Table 4.

Table 4. Robustness

objective	design			prediction		observation			
τ	S_1	S_2	X	Wq_1	Wq_2	$\tau_1$obs	$\tau_2$obs	Wq_1obs	Wq_2obs
99%	67	37	134	0.54	0.22	1	0.07±0.1	0.37±0.01	20.78±1.06
95%	62	34	124	0.71	0.45	0.99	0.03	0.42±0.03	26.99±1.14
90%	60	32	120	0.94	0.76	0.98	0.02	0.45±0.04	30.87±0.1
85%	58	31	116	1.21	1.1	0.94±0.02	0.02	0.65±0.15	34.33±2.08
80%	57	30	114	1.5	1.5	0.93±0.02	0.01	0.65±0.1	35.96±1.14

The fill rate of station 2 is very small because there is constantly a backlog of vehicles at each station. In fact a queue of products will appear at station 2 that will increase untill station 2 calls in enough empty vehicles. The simulation shows that at some point this happens and that the number of products in queue (and thus the waiting time of products) does not grow indefinitely.

In conclusion, we observe that the model suggested offers a good degree of approximation for the fill rate and for the mean waiting time. Moreover the dispatching rule ensures that the system stays in a stable state even if the arrival process is not as expected. This observation gives us an idea for installing a self-adapting system. At regular intervals, we will observe the fill rate for the preceeding period and if the fill rate observed exceeds the fill rate to be achieved, we decrease the dispatching rules limit. If not, we increase the dispatching rules limit.

5.2 Evaluation of the Relevance of our Model

We conclude this evaluation section by showing that our results represent an improvement of a nave approach. Indeed, our model proposes to run the system with much less vehicles than the number of vehicles obtained through other simpler design approaches. Moreover, our dispatching rules guarantee smooth operating conditions even with less vehicles.

A deterministic approach, as the one proposed in Egbelu (1987) and Mahadevan and Narendran (1994), would place all the vehicles at equidistance on the loop and does not take into account the stochastic aspect of the arrival process. In this approach, the number of vehicles needed is $2T\lambda$. For instance, if $\lambda = 0.5$ and $T = 100$ units of time, $X = 100$. The system with 100 vehicles will achieve a very bad level of performance. We need more than 100 vehicles.

Similarly, the anticiped level of performance of a simple queuing approach placing all vehicles at regular time intervals is the mean waiting time of a $M/D/1$ model with a service time of $\frac{2T}{X}$.

$$Wq = \frac{\lambda}{2\mu(\mu - \lambda)}$$

where $\mu = \frac{X}{2T}$.

In our example, to have a mean waiting time less than 1 unit of time, such a simple queuing approach needs 160 vehicles. In our approach, to achieve a Wq of 0.94 units of time, we only need 120 vehicles (see Table 1). So, we see that our model allows one to compute a smaller number of vehicles for the same performance, and proposes adequate dispatching rules to achieve the required performance.

Of course, we have been able to improve the models used to estimate the number of vehicles needed because our transportation network is much simpler. It is the goal of further research to extend the models to more complex transportation settings.

6 Conclusions and Practical Implications

In this paper we address a design problem of a 2-stations AGV system. The objective of this design phase is to achieve a given performance level. The level of performance is the mean waiting time at the stations. We use a combined approach based on three different operation research techniques: inventory management, queuing theory and stochastic processes. Our approach starts by determining a fill rate necessary to achieve this mean waiting time. To achieve a fill rate τ, each station has to maintain a sufficient number of vehicles (at the station or traveling to the station) to cover the demand. To assure this, we determine dispatching rules, according to which a station asks empty vehicles to the other station. To assure that the dispatching rules work correctly, there must be enough vehicles in the system. We propose an

iterative procedure to determine the minimum number of vehicles needed to allow the dispatching rules to work correctly.

We simulate our model and we prove that our model offers better results than simpler deterministic and queuing approaches.

Finally, our model lends itself well to a self-adapting system. Indeed, if the level of performance (Wq) is degraded, the dispatching rules and the number of vehicles can be modified. It is particularly interesting in the case where the mean process arrival varies with time.

Our design model is based on the assumption that there are two stations and that the product is always unloaded at the station following the loading station. One area of future research is to investigate more complex guidepath layouts. First, we want to extend our results to a closed loop of n stations. Secondly, we want to allow the products to be unloaded at any station. In this case, all the vehicles traveling to a station may not be considered to be available at the station when they arrive and we have to know the destination of the product. Finally, in some environments, the guidepath layout might not consist of closed loops. An interesting area for future research would allow the vehicles to use more general routes between stations.

References

Almeida, D. and Kellert, P. (2000) *An analytical queueing network model for flexible manufacturing systems with a discrete handling device and transfer blockings*. The International Journal of flexible manufacturing systems, 12:25-57.

Askin, R. and Standbridge, R. (1993) *Modeling and Analysis of Manufacturing Systems*. Wiley.

Bozer, Y.A. and Kim, J. (1996) *Determining tranfer batch sizes in trip-based material handling systems*. The International Journal of flexible manufacturing systems, 8:313-356.

Egbelu, P.J. (1987) *The use of non-simulation approaches in estimating vehicle requirements in an automated guided vehicle based transport system*. Material flow, 4:17-32.

Gross, D. and Harris, C.(1998) *Fundamentals of Gueueing theory*. John Willey Sons.

Imagine That, I. (1998) *Extend Software*. San Jose.

Johnson, M. and Brandeau, M. (1993) *An Analytical model for design of a multivehicle automated guided vehicle system*. Management Science, 39(12):1477-1489.

Koff, G. (1987) *Automated Guided vehicle systems: Applications, controls and planning*. Material flow, 4:1-6.

Lin, J. (1990) *Microcomputers determine how many AGVs are needed.* Material flow, 22(3):53-56.

Mahadevan, B. and Narendran, T.T. (1990) *Design of an automated guided vehicle based material handling system for a flexible manufacturing system.* International Journal of Production Research, 28:1611-1622.

Mahadevan, B. and Narendran, T.T. (1991) *Estimation of number of AGVs for a FMS - an analytical model.* Industrial Engineering and Management Division.

Mahadevan, B. and Narendran, T.T. (1994) *A hybrid modelling approach to the design of an AGV-based material system for an FMS.* International Journal of Production Research, 32(9):2015-2030.

Malmborg, C. (1990) *Estimation of number of AGVs for an FMS: An analytical model.* International Journal of Production Research, 28(10):1741-1758.

Mantel, R.J. and Landeweerd, H. (1995) *Design and operational control of an AGV system.* International Journal of Production Economics, 41(1-3):257-266.

Maxwell, W.L. and Muckstadt, J.A. (1982) *Design of automated guided vehicle systems.* IIE Transactions, 14(2):114-124.

Newton, D. (1985) *Simulation model calculates how many automated guided vehicles are needed.* Industrial Engin., 68-78.

Rajotia, S., Shanker, K., and Batras, J. (1998) *Determination of optimal AGV fleet size for an FMS.* International Journal of Production Research, 36(5):1177-1198.

Srinivasan, M., Bozer, Y. and Cho, M. (1994) *Trip- based material handling systems: Throughput capacity analysis.* IIE Transactions, 26(1):70-89.

Taghaboni, F. and Tanchoco, J.M.A. (1988) *A LISP-based controller for free-ranging automated guided vehicle systems.* International Journal of Production Research, 26(2):173-188.

Tanchoco, J., Egbleu, P., and Taghaboni, F. (1987) *Determination of the total number of vehicles in an AGV-based material transport system.* Material Flow, 4(12):33-51.

Thonemann, U.W. and Brandeau, M. (1997) *Designing a zoned AGVS with multiple-load capacity.* Operations Research, 45:857-873.

Van der Meer, J. and de Koster, R. (1997) *Centralized versus decentralized control of internal transport, a case study.* B. Fleischmann, J.A.E.E. Van Nunen, L. Grazia Speranza and P. Sthl (Eds), Advances in distribution logistics, Springer (Belrin), 40-420.

Van der Meer, J. and de Koster, R. (1998) *A classification of control systems for internal transport*. Graves et al. (eds), progress in material handling research, Ann arbor, michigan, 633-650.

Van der Meer, J. and de Koster, R. (1999) *Using multiple load vehicles for internal transport with batch arrivals of loads*. 1998 IWDL-springer books, 197-214.

Wysk, R.A., Egbelu, P.J., Chen Zou and Ghosh, B.K. (1987) *Use of spread sheet analysis for evaluating AGV systems*. Material flow, 4:53-64.

Chapter 7

Inventory Control

On-line versus Off-line Control with Multi-load Vehicles

Rene de Koster[1] and J. Robert van der Meer[2]

[1] Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
[2] IG&H Management Consultants, Woerden

Abstract. Computer controlled vehicle-based internal transport systems can be found in many warehouses, manufacturing plants and terminals. The vehicles are usually dispatched on-line, since information on load release times, origins and destinations is usually available only at the last moment. In the theoretical case of off-line control, all load origins, destinations, release instants and transportation times are known in advance. In this case, a mixed integer-programming algorithm can be used to optimize the performance, and heuristics can be used to quickly find good and sometimes optimal solutions.

In this paper, we compare the performance gap between off-line control and on-line dispatching for the case of vehicles with multi-load capacity. The performance is measured as a function of load throughput time, that is, the time a load has to wait when it has been released for transport until it arrives at its destination.

We will show that the performance gap between on-line dispatching and off-line control depends mainly on the throughput. In low throughput environments, vehicles can become idle and park when dispatched on-line. With off-line control, this idle time is used to travel to the next load transportation assignment, hence, compared to on-line dispatching, reducing load waiting times. However, on-line dispatching can outperform off-line control when the actual release times deviate slightly from those that were used to calculate the vehicle routes off-line.

Keywords. Multi-load vehicles, internal transport, warehousing, central control, off-line control, performance, throughput time

1 Introduction

Forklift trucks with vehicle mounted radio terminals and automated guided vehicles, or Guided Vehicles (GVs) in general, are the transportation link for loads between the different areas in transshipment terminals, manufacturing plants, warehouses and distribution centers. To control vehicles for internal transport, several types of control systems can be used. In an earlier paper De Koster and Van der Meer (1998), compared central systems with decentral systems. In the central systems, a central computer keeps track of the vehicles and loads, and assigns one to the other accordingly. In the decentral systems, the GVs drive in loops and trans-

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

port whatever load they encounter first, i.e. the First-Encountered-First-Served rule. The latter rule does not make use of load information such as release times, and it appears that this simple rule is outperformed by the central control systems, which use less vehicles and have smaller load waiting times. In Van der Meer and De Koster (1998), the model for central control systems is extended. Several control rules such as: First-Come-First-Served (FCFS), Nearest-Vehicle-First (NVF), Shortest-Travel-Distance-First (STDF) and Work-List-Dispatching (WLD) are compared. By forecasting the load release times a short time in advance, the authors introduced a virtual release time to signal the vehicles in advance for a pick-up job to be released soon. This extra information in turn reduced load waiting times, but only to a certain extent.

When the forecast time could be extended to a complete day (or daily shift), the vehicles could be routed in such a way as to maximize the performance, i.e. to minimize the load throughput times (the load waiting times plus load travel times). When all move requests at a facility would be known in advance, an efficient schedule could be made off-line to move all requests with a minimum average load throughput time. With the growing use of Electronic Data Interchange (EDI), such scenarios might seem more likely to happen. However, due to last minute updates and unexpected failure of equipment, a stochastic environment is usually the reality. Scheduling vehicles or loads a complete day in advance is therefore near to impossible. In fact, the longer the planning horizon, the less reliable it will be. However, in order to compare off-line control and on-line dispatching we assume in this paper that all information needed for off-line control can be obtained and vehicles can be scheduled optimally. This is done by formulating the situation as a multi-vehicle Pick-up and Delivery Problem with Time Windows, which is solved using Mixed Integer Programming (MIP) for the situation of uni-load vehicles. This exact method using exact information can be seen as an off-line control rule and is compared to the performance of the on-line FCFS and NVF dispatching rules. The idea is to use off-line control performance for uni-load vehicles as a benchmark for on-line performance. We compare the following situations:

- The value of having all information needed for off-line control
- the value of additional vehicles for on-line control
- the value of additional load capacity of vehicles.

Acquiring exact information on loads and load release times is expensive and difficult, if not impossible. Furthermore, solving optimal vehicle schedules requires MIP algorithms that are complex, time consuming and difficult to integrate in vehicle control software. Therefore, perhaps a simple heuristic rule like Insertion (see Solomon (1987) and Solomon & Desrosiers (1988)) is already good enough to be used for off-line vehicle control systems. The latter will also be subject of study in this paper.

Section 2 describes the off-line and on-line control rules used. In section 3 we describe how the vehicles will be dispatched in two different layout environments to investigate the effects of different topologies for different types of dispatching rules. The results will then be discussed in detail in section 4. It will also be shown

that the performance difference between on-line dispatching and off-line control depends mainly on the load throughput and the spread of load-release instances. In low throughput environments, vehicles can become idle and park when dispatched on-line. When controlled off-line, this idle time is used to travel to the next load transportation assignment, hence, compared to on-line dispatching, reducing the average load waiting times and possibly the average throughput times. The paper ends with some concluding remarks.

2 Problem Formulation

As stated before, vehicle-based internal transportation systems can be found in many forms. In each case, loads have to be transported from one location, called the origin, to another, the destination. Loads can be pallets, crates, totes, containers etc.. We consider the general case where loads are released to a transportation system at a certain time i.e. the release time, and need to be transported to a destination. The object is to minimize the average time between the release time of the load and the drop-off time of the load at its destination. This time, the average load throughput time, defines the performance of the control rule for the vehicles. The smaller the average load throughput time, the better the performance. In other words, loads have to be picked up after their release time (start of time window) by one of the vehicles and brought to their destination in such a way that the average load throughput time is minimal. It is desirable to keep the average load waiting time to a minimum in order to service waiting trucks at the shipping lanes, release new space at small output buffers, quickly transport perishable products to cooled areas, service other handling equipment, etc.. The transportation time can usually not be influenced much, except in the case of multi-load vehicle control. When a load transport is combined with another load transport, the first load has some extra travel time. This is the main reason why we define the performance as the sum of the load waiting times and load travel times.

In the off-line case, where all transport jobs, including release times, are known in advance, the problem can be modeled as a multi-vehicle pick-up and delivery problem with time windows (m-PDPTW) where the objective is to minimize the load throughput time. The following sections describe the off-line control rules used. The vehicle routes can be optimized with mixed integer programming algorithms. The m-PDPTW is NP-hard and the algorithms to solve this problem optimally are very time and memory consuming. We therefore also describe a heuristic to solve larger problem instances.

2.1 Off-line Control Rules

The following sections describe the off-line control rules used. Because all information is known beforehand, the vehicle routes can be optimized with Mixed In-

teger Programming (MIP) algorithms. However, in practice, these algorithms can be very time and memory consuming. We therefore also describe an Insertion heuristic and use it to solve the same and larger problem instances.

2.1.1 The Multi-Vehicle Pick-up and Delivery Problem with Time Windows

An extensive discussion on the PDPTW is given by Dumas *et al.* (1991). In their model the vehicles must pick up a load at the load origin between the start and end of the pick-up time window and deliver the load at its destination between the start and end of the destination time window. However, the time window formulation for the study in this paper is different. The release time of the load defines the start of the pick-up time window, so the loads can be picked up any time after that and must be delivered directly. Furthermore, to keep the problem computationally tractable, only uni-load vehicles are used for our MIP algorithm. This is also referred to as a full-truck load problem, see Savelsbergh and Sol (1995).

A full-truck load pick-up and delivery problem can be formulated as a Traveling Salesman Problem (TSP) by representing a transportation job as a single job-node (instead of an origin and destination location-node) in which the travel time from job i to job j (t'_{ij}) equals the travel time from the origin of job i to the destination of i ($t_{i^+i^-}$) plus the travel time from the destination of i to the origin of j ($t_{i^-j^+}$). So $t'_{ij} = t_{i^+i^-} + t_{i^-j^+}$. This origin to origin formulation would cause problems for the time window constraints for destination locations. However, in the formulation of this paper, there are only start time constraints at the origins. The PDPTW in this special case can therefore be formulated as a TSP with time windows (TSPTW) with the objective to minimize the load throughput time, which is the sum of all load waiting times plus load travel times.

With multi-load vehicles the load travel times would not be unique since a load's travel time can increase when another load is picked up and dropped off first by the loaded vehicle. In the case of uni-load vehicles, the load transportation times are constant and the objective reduces to the minimization of the load waiting time (we will add the loaded trip times at the end). This in turn is also referred to a Traveling Repairman Problem with time windows (TRPTW) (see also Ball et al. (1995)). The formulation in the next section has no restriction for the end-time of the time window; this is the main difference with other TRPTW formulations found in literature and is discussed in more detail in the next section.

2.1.2 The Traveling Repairman Problem with Time Windows

We give the formulation for the TRPTW involving a single depot (which is represented by a node where the vehicles start from to serve their first job, and return to after completing their last job) and a homogeneous fleet of vehicles for the models studied in this paper. The notation used is listed in Table 1, the mathematical formulation is listed in Formulation 1.

Table 1. Notation for the TRPTW

Index sets	
N	set of nodes $\{0, ..., n + 1\}$ for the vehicle network, indexed by i and j
P	set of nodes $\{1, ..., n\}$ other than the depot nodes
V	set of vehicles $\{1, ..., \|V\|\}$ to be routed where $\|V\|$ is the number of vehicles, indexed by v
Parameters	
n	number of load transportation jobs $\|P\|$, associate to job i a node i
r_i	release time of the load at node i, (which defines the start of the pick-up time window)
t_{ij}	travel distance/time from i to j for each distinct i, j in N (that is from the origin of load i to the origin of load j)
Variables	
x^v_{ij}	binary flow variables which equal 1 if vehicle v travels from node i to node j and zero otherwise, $v \in V$, $i, j \in N$
D_i	time at which service at node i begins, $i \in P$
D^v_0	time vehicle v leaves the start depot (node 0), $v \in V$
D^v_{n+1}	time vehicle v returns to the end depot (node $n + 1$), $v \in V$

As vehicles travel on the network transporting loads form one location to another, some locations are visited more than once. However, a (job-)node is associated to each transportation job in order to assign a unique service or departure-time to each job. Therefore different nodes may refer to the same physical location at which a transport request was placed. Since each vehicle starts and ends its route at the depot, the depot would be associated with several service-times. However, a variable can only be associated with one value. Therefore extra dummy service-time variables (D^v_0, D^v_{n+1} were $v \in V$) are introduced for the depots which all refer to the same physical depot location.

$$Min \sum_{i \in P} (D_i - r_i) \quad (1)$$

Subject to

$$\sum_{v \in V} \sum_{j \in N} x_{ij}^v = 1 \quad \forall i \in P \quad (2)$$

$$\sum_{j \in N} x_{ij}^v - \sum_{j \in N} x_{ji}^v = 0 \quad \forall i \in P, \forall v \in V \quad (3)$$

$$\sum_{j \in P} x_{0j}^v = 1 \quad \forall v \in V \quad (4)$$

$$\sum_{i \in P} x_{in+1}^v = 1 \quad \forall v \in V \quad (5)$$

$$x_{ij}^v = 1 \Rightarrow D_i + t_{ij} \leq D_j \quad \forall i, j \in P, \forall v \in V \quad (6)$$

$$x_{0j}^v = 1 \Rightarrow D_0^v + t_{0j} \leq D_j \quad \forall j \in P, \forall v \in V \quad (7)$$

$$x_{in+1}^v = 1 \Rightarrow D_i + t_{in+1} \leq D_{n+1}^v \quad \forall i \in P, \forall v \in V \quad (8)$$

$$D_i \geq r_i \quad \forall i \in P \quad (9)$$

$$D_0^v = 0 \quad \forall v \in V \quad (10)$$

$$D_{n+1}^v \geq 0 \quad \forall v \in V \quad (11)$$

$$\sum_{v \in V} \sum_{j \in N} x_{n+1j}^v = 0 \quad (12)$$

$$\sum_{v \in V} \sum_{i \in N} x_{i0}^v = 0 \quad (13)$$

$$x_{ij}^v \quad binary \quad \forall i, j \in N, \forall v \in V \quad (14)$$

Formulation 1: **The mathematical formulation of the TRPTW**

We seek to minimize the sum of the load waiting time (see equation (1)), i.e. the sum of differences between the departure time D_i of a vehicle at node i, and the release time/earliest possible pick-up time r_i of the load at that node. The corresponding objective implicitly minimizes the *average* load waiting time as well. And when the loaded trip times are added, it also minimizes the average load throughput time. If a vehicle arrives at a node before the load is released, the vehicle must wait. Constraints (2)-(5) and (12)-(13) form a multi-commodity flow formulation, in which constraint (2) ensures that all nodes are visited exactly once. Constraint (3) in turn ensures that a vehicle arriving at a node will also leave that node. Furthermore, vehicles must leave the starting node (constraint (4)), and constraint (13) makes sure that no vehicle can return to the starting node. Constraints (5) and (12) make sure vehicles arrive at the end node and never leave from the end depot respectively.

Next, constraints (6)-(8) describe the compatibility requirements between routes and schedules, while constraints (9)-(11) are the time window constraints.

Constraint (9) defines the start of the pick-up time window, since vehicles can come any time after the release time; there is no constraint for the end time.

Constraints (6)-(8) in Formulation 1 are not linear, but can be rewritten in an equivalent linear form using a large constant M:

$$D_i + t_{ij} - D_j \le M\left(1 - x_{ij}^v\right) \quad \forall i, j \in P, \forall v \in V \qquad (6`)$$

$$D_0^v + t_{0j} - D_j \le M\left(1 - x_{0j}^v\right) \quad \forall j \in P, \forall v \in V \qquad (7`)$$

$$D_i + t_{in+1} - D_{n+1}^v \le M\left(1 - x_{in+1}^v\right) \quad \forall i \in P, \forall v \in V \qquad (8`)$$

Constraints (6`)-(8`) also impose increasing times at the nodes of the route. Thus, eliminating possible cycles. These constraints are in fact a generalization of the subtour elimination constraints proposed by Miller et al. (1960).

Since a node has a pick-up time window with a start time only, a vehicle can arrive any time after that. This will result in a large number of possible routes. By introducing an end for the pick-up time window like constraint (9`) below, some routes are eliminated and thereby the speed of finding the optimum is increased.

$$D_i \le r_i + C \qquad \forall i \in P \qquad (9`)$$

In this case, a constant C is used to form a time window of length C in which a vehicle should pick up the load at node i. However, setting the pick-up time window too narrow will lead to a suboptimal solution (possibly even an infeasible one if all feasible routes are eliminated, C should then be increased). The (suboptimal) value of the objective function of this 'previous run solution' can be used to set a new end time for the pick-up time window. Adding the 'previous run solution' to all load release times will create new end times for the pick-up time window (see constraint (9``) next page). This will lead to the optimal value when the MIP is run again, since the optimal value will always be smaller (or equal) than the time windows created with a 'suboptimal' answer.

More formally, since

$$\left[\sum_{i \in P}\left(D_i - r_i\right)\right]^{optimal\ solution} \le \left[\sum_{i \in P}\left(D_i - r_i\right)\right]^{previous\ run}$$

it follows that all individual waiting times of the optimal solution are smaller than the sum of the waiting times, and

$$(D_i - r_i)^{optimalsolution} \leq \left\lceil \sum_{i \in P} (D_i - r_i) \right\rceil^{optimalsolution} \leq \left\lceil \sum_{i \in P} (D_i - r_i) \right\rceil^{previous\,run} \qquad \forall i \in P$$

which leads to

$$D_i \leq r_i + \left\lceil \sum_{i \in P} (D_i - r_i) \right\rceil^{previous\ run} \qquad \forall i \in P \qquad (9``)$$

This means that for the optimal value, the MIP should be run again with constraint (9`) replaced with (9``). However, the second run is only necessary when the solution of the first run is larger than C. Note that the solution of a heuristic algorithm (such as Insertion) can also be used to estimate C. With such an (over)estimate of C, only one run is necessary to obtain the optimal value. However, the running time can be relatively high since the bound can be rather weak.

2.1.3 The Insertion Rule

When using CPLEX to solve the MIP model, memory problems (over 125 MB of RAM was available) and long running times (on an IBM/RS6000 model 370) were soon encountered for relatively small problems (see also Section 4.1). To decrease the running time, increase the problem size and increase the chance of practical implementation, we also analyzed the results with an Insertion heuristic. Insertion heuristics have been studied for a variety of vehicle-routing problems (see Solomon (1987)), dial-a-ride problems (see Jaw et al. (1986)) and traveling-salesman problems (see Gendreau et al. (1992)). Insertion heuristics have shown very promising results in these studies. For off-line vehicle-control based on the traveling repairman problem of Section 2.1.2, we will describe an Insertion type heuristic and compare the results with the optimal results for small problem sizes. The Insertion heuristic will then also be used for larger problem instances.

The pseudocode of the Insertion algorithm used to construct the vehicle routes off-line is presented in Algorithm 1. After sorting the jobs in increasing order in terms of the release time, the position with the cheapest insertion cost of the job is calculated for each job. This is the minimal extra waiting time needed to add job i to a vehicle route v. Since the number of candidate positions is at most n, and the number of jobs considered to one of the positions is at most n, an algorithm that enumerates all jobs for all candidate positions (the Insertion heuristic) will have a time complexity of $O(n^2)$. The algorithm is actually carried out twice in case a job can be inserted in different vehicle routes with the same costs (ties). The first time the data about the possible insertion position is not updated in case of a tie, (so the job is assigned to the first route encountered with that solution). The second time the data about the possible insertion position is also updated when a tie is encountered, (so the job is assigned to the last route encountered with that solution).

```
Perform jobs to job-nodes transformation;
Construct the node-list by sorting tasks on increasing release times;
for v := 1 to |V| do
Initialize vehicle route v with depot D^v_0;
for i := 1 to n do {
        for v := 1 to |V| do {
                for j := 1 to (nr. nodes of vehicle route v) + 1 do {
                Temporarily insert node i at position j of route v;
                Recalculate the sum of differences between departure and release times
            (waiting times) for all inserted nodes so far; }}
        Insert node i at position j of vehicle route v for which the total sum of waiting times
    was minimal; }
Report total waiting time;
```

Algorithm 1: **Simplified algorithm of the Insertion heuristic in pseudocode**

The next situation of four locations (including the depot), three jobs and two vehicles is an example of the Insertion procedure. The travel times between the four locations for this example are shown in Table 2. Table 3 gives the jobs to job-nodes transformation (origin-to-origin) and Table 4 the corresponding node to node travel times. For example, the travel time from node 2 to node 3 in Table 4 is the travel time from location 1 to location 3 (the third job row in Table 3), which equals 20 (see Table 2) plus the travel time from (destination) location 3 to (origin) location 1.

Table 2. Travel times between locations

Location	Depot	1	2	3
Depot	0	10	10	20
1	10	0	10	20
2	10	10	0	10
3	20	20	10	0

Table 3. Jobs to job-nodes transformation

Load release time (r_i)	Job	Node (i)
0	Depart depot	0
9	From location 3 to location 1	1
15	From location 1 to location 3	2
19	From location 1 to location 2	3

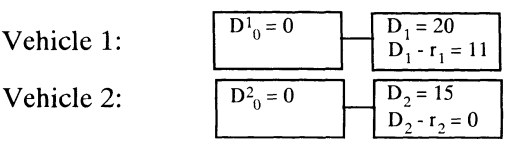


Fig. 1. Load to vehicle assignments after two insertions

Table 4. Travel times between nodes

Node	0	1	2	3
0	-	20	10	10
1	-	-	20	20
2	-	20	-	40
3	-	20	20	-

Figure 1and Figure 2 represent how the vehicle routes are constructed using the Insertion heuristic with the notation of the TRPTW for the departure and release times, the release times of Table 3 and the node to node travel times of Table 4. After two Insertion steps, each vehicle has one job and the sum of the load waiting times is as small as possible (11), as shown in Figure 1. In the next step, the third job is inserted in the most favorable position of the route for one of the vehicles, giving the four alternatives shown in Figure 2.

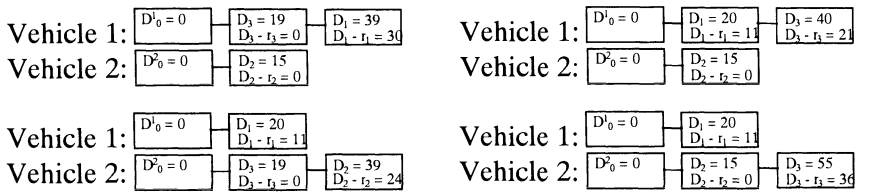


Fig. 2. Load to vehicle assignment possibilities when inserting job 3

In this case, the first alternative leads to the smallest total waiting time (30), and job 3 is inserted at the beginning of the route (after leaving the depot) of vehicle 1. In the case of more jobs and vehicles, the algorithm proceeds in a similar fashion checking all possibilities until all jobs are assigned to a vehicle.

Although this algorithm will not guarantee an optimal route, we can still use the value of off-line control systems when using a simple heuristic by demonstrating that the solutions are sufficiently close to the optimum. These solutions can still be further improved by using more advanced heuristics such as those using column generation techniques (see Dumas et al, 1991). The latter are beyond the scope of this study and will not be discussed.

2.2 On-line Dispatching Rules

In this paper, two on-line control rules are compared where loads and vehicles can only be assigned after the release of the loads (the moment they can be transported). The first rule described is a distance-based rule, which has been studied in earlier papers of De Koster and Van der Meer (1998), and Van der Meer and De

Koster (1998) and has proven to be very effective. We also include a time-based rule studied by Boozer *et al.* (1994), De Koster and Van der Meer (1998), Srinivasan *et al.* (1994) and Van der Meer and De Koster (1998). Both rules use the same data used with off-line control. In this case however, the information of the jobs is made available to the vehicles at the release time of the loads.

Nearest-Vehicle-First

Under this rule, the load or workcenter has the dispatching initiative. When a load is released at a workcenter, the workcenter places a move request. The shortest distance along the traveling paths to every available (idle and motionless) vehicle is then calculated. The idle vehicle, whose travel distance to the load is the shortest, will be awoken to be dispatched. On the other hand, when a vehicle becomes idle, it searches for the closest waiting load in the system, i.e., at that point the dispatching initiative is at the vehicle and the rule used is similar to shortest-travel-distance-first (STDF). If there are no vehicle requests for loads in the system, the (empty) vehicles will park at their current locations and become idle until a request becomes available.

Modified First-Come-First-Served

The FCFS rule is a vehicle-initiated dispatching rule. A vehicle delivering a load at the input queue of a station first inspects the output queue of that station. The vehicle is then assigned to the oldest request (longest waiting load) at that station if one or more loads is found. However, if the output queue of that station is empty, the vehicle serves the oldest request in the entire system. If there are no move requests in the system at all, the vehicle will park at that location and becomes idle until a move request becomes available.

Multi-load vehicle dispatching

Multi-load vehicle dispatching is based on the concept of closest task. Therefore, a multi-load vehicle picks up as many loads as it can carry from its current location before moving away. When the vehicle moves, it either delivers one of its loads or picks up another load if it has remaining capacity. The vehicle only looks for additional loads to pick up that are closer in distance than the closest destination of its onboard loads. If the vehicle goes to deliver a load, it always goes to the closest among the destinations of its onboard loads.

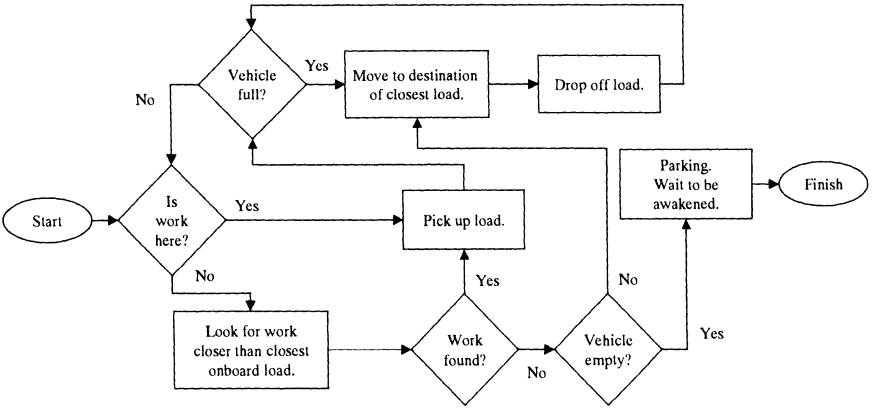


Fig. 3. Vehicle dispatching behavior

The concept of closest task for multi-vehicle dispatching applies to the previously described NVF and FCFS dispatching rules. The flowchart of Figure 3 shows the decisions made during vehicle-initiated dispatching. When a vehicle drops off a load, the vehicle continues by checking for (additional) work. When vehicles are parked when a load is released in the system, the (idle) vehicles are awoken which then check for (additional) work.

Figure 4 shows how vehicle behavior is affected with load-initiated rules. If a load is released and no (idle) vehicle is found with remaining capacity, the dispatching initiative is passed to the vehicles.

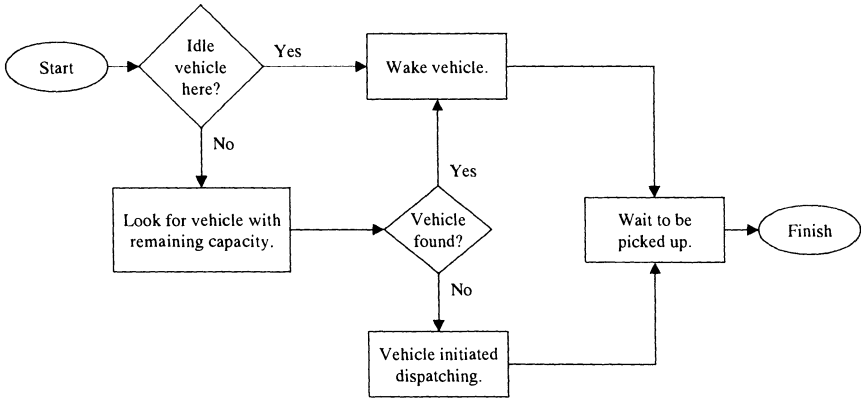


Fig. 4. How loads affect vehicle behavior

So the performance of the NVF and modified FCFS rules is mainly characterized by the dispatching rule triggered by the first onboard load when multi-load vehicles are used. In this study the capacity of the vehicles is at most two; i.e. dual-load vehicles.

3 The Model

3.1 The U-layout and I-layout Environments

Figure 5 gives a representation of an I-layout and U-layout transportation environment respectively; two common warehouse layouts found in practice. The dashed lines represent the contours of the building. The solid lines represent the network on which the guided vehicles travel. The vehicles are stored in the vehicle depot and also start and end their daily tasks at the depot. The other nodes on the vehicle path represent different locations (the origins and destinations of loads) which the vehicles visit to serve transportation requests. The numbers beside the paths represent the distance units between the nodes when that path is followed. These numbers can also be seen as time units since the vehicles travel with constant unit speed, i.e. one distance unit per time unit.

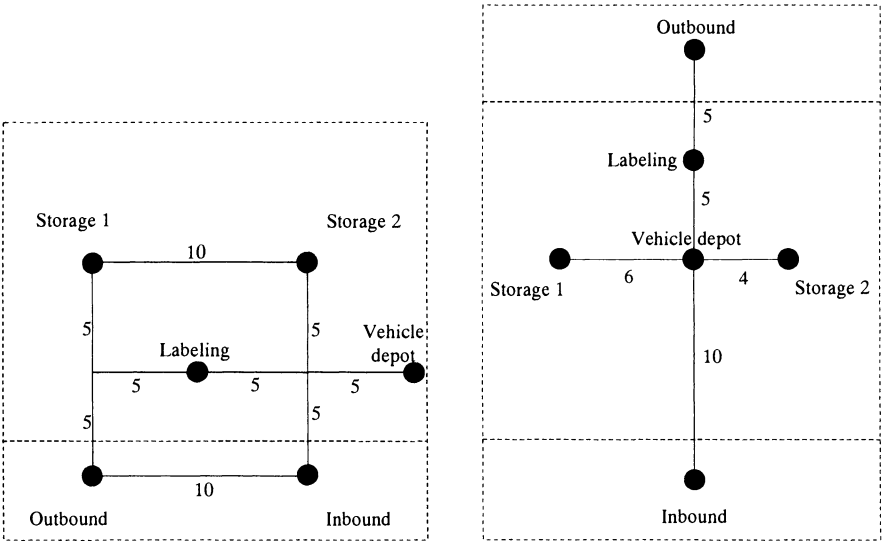


Fig. 5. Representation of the U-layout (left) and I-layout (right) warehouses

The design of the facility is mainly dependent on the nature of activities being performed inside the facility and the access to outside transportation facilities, see Tompkins et al. (1996). If both receiving and shipping occur simultaneously, then close supervision is required to ensure that received goods and goods to be shipped are not mixed.

If storage is one of the main functions of the warehouse, then both the receiving (Inbound) and shipping (Outbound) lanes are usually at one side of the building. The result is a so-called U-layout warehouse with a rectangular shape (see Figure 5). In this way it is possible to partly utilize the same docks, personnel and han-

dling equipment for shipping and receiving operations. The storage modules are at the other side of the building and the stations with for example, added value logistics (VAL), in this case a labeling station, in the middle of the warehouse.

The I-layout warehouse (see Figure 5) is an example warehouse commonly used in situations where transshipment is the most important process and storage is less important. Loads are received at one end and leave at the other end. Hence, the receiving (Inbound) and shipping (Outbound) lanes are at opposite ends of the warehouse, and all other stations more or less in the middle.

An extra advantage of the U-layout is the greater possibility for double-plays (combining inbound trips with outbound trips) since the Outbound and Inbound areas are relatively closer to each other. This means that vehicles may be better utilized since empty travel times decrease (and load waiting times possibly decrease).

In this case, the advantage of the I-layout is the smaller transport distances for stored material from Storage 1 and 2 towards the Outbound lanes. The disadvantage is the greater distance between the Inbound and Outbound areas, which slightly increases the average distance between any location to any other location. Observe that the I-layout is less symmetric in distances than the U-layout. This has been done on purpose, in order to investigate whether symmetry has an effect on the performance of certain dispatching rules. Intuitively, one can imagine that a distance-based dispatching rule works better if there are differences in the travel distances, like in a non-symmetrical environment. The vehicle paths for both warehouses are bi-directional and vehicles may pass each other if necessary. The pick up and set down times of loads are negligible and idle vehicles park at their current location.

In the case of the example warehouses, Inbound loads arrive at the Inbound area and are transported to Storage 1 or Storage 2. At Storage 1 and Storage 2, loads that need to be transported are sent to the Labeling area. From the Labeling area Outbound loads are transported to the Outbound area. In both U-layout and I-layout situations, the average inbound travel time is the same. For the U-layout, the travel time is either 10 or 20 units; this means 15 units on average. For the I-layout, the travel time is either 16 or 14 units; this also means 15 units on average. The Outbound loads first go through the Labeling area. In the U-layout this means that those loads always travel 20 units. As mentioned before, the Outbound loads of the I-layout have a travel time advantage. In this case the Outbound loads travel 15 units on average.

So there are three classes with a total of 5 job types:

Class 1: Inbound
1) Inbound to Storage 1 (travel time: 20 time units for U-layout and 16 for I-layout);
2) Inbound to Storage 2 (travel time: 10 time units for U-layout and 14 for I-layout);

Class 2: Labeling
3) Storage 1 to Labeling (travel time: 10 time units for U-layout and 11 for I-layout);
4) Storage 2 to Labeling (travel time: 10 time units for U-layout and 9 for I-layout);

Class 3: Outbound
5) Labeling to Outbound (travel time: 10 time units for U-layout and 5 for I-layout).

In general, two vehicles are used for the transportation jobs. However, in case of on-line dispatching, more vehicles are needed to transport all loads in the given time period. The jobs are generated such that three different daily shifts are constructed with different throughput characteristics, as described in the following sections.

3.3 Random Shifts

To keep the MIP problem computationally tractable, the number of jobs could not exceed 12 (see also Section 3.6.1). Since there are three classes, the idea is to generate four jobs of each class on average. Using a uniform distribution, job types are generated at random (where the Outbound job type is weighed twice). It is then easy to see that the (daily) shift of 12 jobs on the U-layout has a total loaded trip time of 140 time units (see Table 5) on average, or 70 time units per vehicle on average. When the empty trip time is estimated at 80% of the loaded trip time, or 56 time units per vehicle, the total trip time will be 126 time units.

Table 5. Average total loaded travel time units per layout

	U-lay-out	I-lay-out
Job type	Loaded travel time units	Loaded travel time units
Inbound	2 520 + 2 510	2 516 + 2 514
Labeling	4 510	2 511 + 2 59
Outbound	4 510	4 55
Total	140	120

Although similar calculations for the I-layout result in an average total trip time of 108 time units per vehicle (see Table 5), the same data (transport jobs and load release times) generated with the calculations of the U-layout is used for both layouts. Jobs for both layouts are therefore generated between 0 to 126 time units (the daily shift) from a uniform distribution. These jobs are then assigned to the vehicles according to the dispatching rule used. Observe that one vehicle can have more job assignments than another. The average load throughput times are calculated over a total of 10 different generated shifts.

3.4 Structured Shifts

In the case of Random shifts, the 12 jobs are uniformly generated over a period of 126 time units. With Structured shifts, one-third of the jobs are uniformly generated over the first 40% of the total shift time and consist (only) of Inbound jobs. The last 40% of the shift consist of Outbound jobs (also one-third of the total amount of jobs), and the middle 40% (so there is an overlap of 10% on each side) consists Labeling jobs, see also Figure 6.

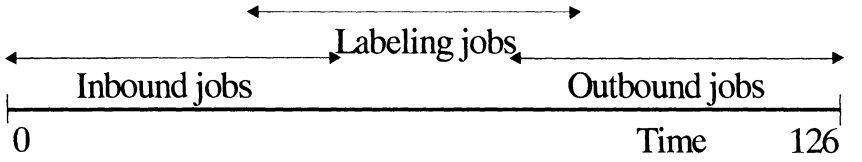


Fig. 6. Structured dispersion of jobs over the shift

The Structured shifts comes from the idea that in many warehouse and manufacturing situations; inbound jobs precede processing jobs, which in turn precede outbound jobs during a day.

3.5 High Throughput Shifts

In order to investigate the dependency of the performance gap between on-line dispatching and off-line control on throughput and vehicle utilization, we increased the throughput level. In many environments, peaks of high workloads can be observed. When the jobs are generated in a shorter time period, the probability of on-line dispatched vehicles *waiting* for a transport assignment is reduced, similar to peak behavior. So by reducing vehicle idle times, the extra load waiting times will be reduced and we can expect that the performance of on-line dispatching and off-line control will be closer together. This is the main idea of High throughput shifts.

The jobs are uniformly generated in a structured shift as shown in Figure 6. The number of jobs, however, is increased to 60 (20 Inbound, 20 Labeling and 20 Outbound). In this case, the jobs are generated over a period equal to the average loaded trip time of the U-layout (so 10*20 + 50*10 = 700 time units) plus an extra 20% to account for (some) of the empty trip time; in total 420 time units per vehicle. Again, although similar calculations for the I-layout result in an average total trip time of 360 time units per vehicle, the data generated with the calculations of the U-layout is used for both layouts. The Insertion heuristic is used as the off-line control rule since exact off-line control appeared to be intractable for this situation.

4 Results

The discussion of the results will start by presenting the performance gap, i.e. the differences in expected load throughput times, between off-line and on-line controlled uni-load guided vehicles. It will also be shown to what extent the on-line controlled vehicle fleet has to increase to approximate off-line performance. Next, we present the effects of using dual-load vehicles with on-line control and compare the results with increasing the fleet size with uni-load and dual-load vehicles.

4.1 Varying the Number of Vehicles

Table 6 gives an overview of the average throughput times (see 'Average') and standard deviation (see 'St. dev.') of the throughput times for 10 runs for both off-line control and on-line dispatching for both layout types in the Random shifts situation. For the off-line rules, 'Optimal' refers to the optimal solution from the TRPTW, i.e. the minimum load throughput times possible, when two guided vehicles (GVs) are used. In case of 12 jobs, the computation times for the Optimal solution solving the TRPTW with CPLEX, varied from two minutes to several hours on an IBM/RS6000 class computer, and in some instances up to 120 Mb of memory was required for the branch and bound tree. The throughput times in the 'Insertion' column represent the throughput times obtained when two vehicles are routed with the Insertion heuristic (less than one second computation time). This leads to the optimal solution in several instances; overall it deviates about 4% and 5% from the optimal value (see 'Deviation' in Table 6)for the U-layout and I-layout respectively. When on-line dispatching rules are used, the loads have to wait about twice as long to be transported with the same number of vehicles and the average throughput time is more than 50% higher in the I-layout case. Notice that the average throughput times are smaller in the I-layout environment. This is due to the travel time advantage for Outbound loads.

To bring the average throughput times with on-line dispatching within 10% of the optimal solution, the fleet size had to be doubled to four vehicles.

Table 6. Average load throughput times with Random shifts (10 runs consisting of 12 jobs)

	Off-line control		On-line control: NVF			On-line control: FCFS		
	Optimal	Insertion	2 GVs	3 GVs	4 GVs	2 GVs	3 GVs	4 GVs
U-lay-out								
Average	235.2	244.1	328.2	260.7	245.1	321.5	261.4	248.0
St. dev.	20.7	22.9	41.5	24.4	22.9	47.5	26.4	18.3
Deviation	-	4 %	40 %	11 %	4 %	37 %	11 %	5 %
I-lay-out								
Average	195.0	204.6	300.7	235.2	210.3	304.4	230.5	219.6
St. dev	44.6	45.1	44.3	33.7	27.0	38.2	45.5	26.9
Deviation	-	5 %	54 %	21 %	8 %	56 %	18 %	13 %

Even when the fleet size is quadrupled (not in table), the optimal off-line rule outperforms the on-line rules on average. This is due to the fact that off-line controlled vehicles can use idle time to move closer to the next task, hence reducing load waiting time. In the Random shifts case, the vehicle idle time with on-line control appears to be about 15% (not shown in the table), this means that there is some slack in the system which could be used to route the vehicles to the next assignment and reduce the load waiting times.

Table 7. Average load throughput times with Structured shifts (10 runs consisting of 12 jobs)

	Off-line control		On-line control: NVF			On-line control: FCFS		
	Optimal	Insertion	2 GVs	3 GVs	4 GVs	2 GVs	3 GVs	4 GVs
U-lay-out								
Average	215.6	218.5	274.8	221.8	211.0	277.0	214.6	193.0
St. dev.	12.8	12.8	20.3	19.6	15.6	19.2	16.5	11.0
Deviation	-	1 %	27 %	3 %	-2 %	28 %	0 %	-10 %
I-lay-out								
Average	198.4	201.1	261.7	197.9	173.1	262.8	197.8	175.3
St. dev.	23.6	25.2	19.9	17.4	10.8	23.6	16.3	13.7
Deviation	-	1 %	32 %	0 %	-13 %	32 %	0 %	-12 %

Next, the case of Structured shifts is studied. Because of the design of the warehouse and the overlapping periods within the shifts, a combination of dropping off an Inbound load and picking up a load for Labeling or the combination Labeling and Outbound loads, leads to smaller waiting times and thereby smaller load throughput times with the uni-load vehicles. This can be seen by comparing the values of Table 6 with those of Table 7.

For two vehicles, the average load throughput times with online dispatching are about 30% higher than the optimum. The difference between Insertion and the Optimal value is about 1% and the fleet only needs 50% extra vehicles instead of twice as many to approximate Off-line performance within 10%. We can conclude that Structured shifts leads to a better performance for both off-line and on-line control than when the jobs are Random in a shift. Although the differences in the average load throughput times between NVF and FCFS are small, it seems that NVF is a little more favorable in the I-layout environment and FCFS in the U-layout environment (in both cases two out of three times on average, see Table 7). Since the U-layout is rather symmetrical in travel times, the dispatching decisions with a time-based rule turn out to be more favorable. In the less symmetrical I-layout, the dispatching decisions can be made based on different travel distances and the distance-based rule turns out to be more favorable.

In the next experiment, the transport request intensity is increased. This means that there will be less idle time for the vehicles, which will reduce the performance gap between off-line control and on-line dispatching. A total of 60 loads are generated in a time frame that has a length of 1.2 times the load transport time. This

will be done in a similar manner, as was the case for 12 loads. The extra 20% is added to account for (some) empty trip time.

We continued the study without calculating the Optimal performance with TRPTW since this led to high running times and computer memory problems. Since the Insertion heuristic leads to very satisfactory results (see the previous part of this section) in a very simple and quick way, we will continue to use Insertion for the off-line control.

Table 8. Average load throughput times with High throughput shifts (10 runs with 60 jobs)

	Off-line control: Insertion		On-line control: NVF		On-line control: FCFS	
	2 GVs	3 GVs	2 GVs	3 GVs	2 GVs	3 GVs
U-lay-out						
Average	6422.8	2006.3	6907.3	2355.1	6805.0	2417.8
St. dev.	629.0	361.3	700.0	382.4	582.0	412.1
Deviation	-	-	8 %	17 %	6 %	21 %
I-lay-out						
Average	6073.2	2095.2	6518.7	2592.3	7265.9	2586.6
St. dev.	730.0	847.8	710.2	834.1	757.3	791.3
Deviation	-	-	7 %	24 %	20 %	23 %

The results in Table 8 show that the deviation in load throughput times between on-line dispatching and off-line control is smaller when extra waiting time is eliminated by removing the slack in the system. For the 2-vehicle situation, the load throughput times with on-line dispatching are about 8% and 7% higher, and 17% and 24% for the 3-vehicle situation for the U-layout and I-layout respectively. Considering that the load throughput times with Insertion were up to 5% above the optimum (see the results for Random shifts), we expect that the deviation with the optimal performance is still reasonable.

4.2 Varying the Capacity of Vehicles

The alternative to increasing the number of vehicles is to increase the vehicle capacity. In general, two uni-load vehicles are more expensive than one dual-load vehicle, while the number of loads that can be transported simultaneously is the same.

With doubling the fleet size as in the previous section, we risk congestion, etc. Instead, we would like to see the effects of doubling the vehicle capacity, although the control of dual-load vehicles is more complex than the control of uni-load vehicles as explained earlier. The dual-load vehicles are used with NVF and FCFS dispatching only and are also compared with the performance (average load throughput times) of off-line controlled (Insertion) uni-load vehicles. From the results in Table 9, we see that increasing the vehicle capacity leads to a limited increase in performance and diminishes as the fleet size increases. The performance

of two dual-load vehicles, (when four loads can be transported simultaneously), is worse than the performance of three uni-load vehicles. In fact, four uni-load vehicles (see Table 6) outperform three dual-load vehicles (see Table 9).

Table 9. Average load throughput times for Random shifts (12 jobs)

	Off-line control		On-line control: NVF				On-line control: FCFS			
	Inser-tion: 2 GVs	Inser tion: 3 GVs	2 GVs cap. 1	2 GVs cap. 2	3 GVs cap. 1	3 GVs cap. 2	2 GVs cap. 1	2 GVs cap. 2	3 GVs cap. 1	3 GVs cap. 2
U-lay-out										
Aver age	244.1	176.1	328.2	300.7	260.7	255.0	321.5	298.2	261.4	256.8
St. dev.	22.9	13.9	41.5	38.6	24.4	24.3	47.5	36.5	26.4	29.2
Devia tion	-	-	34 %	23 %	48 %	45 %	32 %	22 %	48 %	46 %
I-lay-out										
Aver age	204.6	141.1	300.7	272.1	235.2	233.7	304.4	278.6	230.5	240.0
St. dev.	45.1	15.4	44.3	43.7	33.7	30.2	38.2	43.4	45.5	33.1
Devia tion	-	-	47 %	33 %	67 %	66 %	49 %	36 %	63 %	70 %

In Table 9 we can also see the negative effects of dual-load vehicles on the load throughput time. Although the load waiting time can reduce when dual-load vehicles are used, the load transportation time can increase. Load transportation times can increase since loads do not have to be delivered immediately after being picked up. Certain loads can remain on the vehicle while other loads are serviced with the remaining vehicle capacity. The result is that the sum of the load transportation time and load waiting time (i.e. defined as the load throughput time), can then also increase. This phenomenon can be seen in the I-layout environment when three vehicles are dispatched with the FCFS rule. When the vehicle capacity increases, the average load throughput time increases from 230.5 time units to 240 time units.

Since transportation jobs are more structured in the Structured shifts, the opportunity for combining transportation jobs with dual-load vehicles increases. It can be seen in Table 10 that the deviations are more favorable compared to Random shifts in Table 9. The load throughput time can still increase as can be seen in the I-layout environment when three vehicles are dispatched with the FCFS rule, but to a less extent compared to Random shifts.

Table 10. Average load throughput times for Structured shifts (12 jobs)

	Off-line control		On-line control: NVF				On-line control: FCFS			
	Insertion: 2 GVs	Insertion: 3 GVs	2 GVs cap. 1	2 GVs cap. 2	3 GVs cap. 1	3 GVs cap. 2	2 GVs cap. 1	2 GVs cap. 2	3 GVs cap. 1	3 GVs cap. 2
U-lay-out										
Average	218.5	164.1	274.8	256.9	221.8	219.2	277.0	256.9	214.6	210.3
St. dev.	12.8	10.8	20.3	21.9	19.6	19.9	19.2	21.9	16.5	13.1
Deviation	-	-	26 %	18 %	35 %	34 %	27 %	18 %	31 %	28 %
I-lay-out										
Average	201.1	150.1	261.7	245.1	197.9	197.1	262.8	244.1	197.6	197.7
St. dev.	25.2	8.9	19.9	14.4	17.4	18.5	23.6	14.2	16.7	18.5
Deviation	-	-	30 %	22 %	32 %	31 %	31 %	21 %	32 %	32 %

We also see that increasing the vehicle capacity leads to a smaller decrease in average throughput times compared to Random shifts, and the decrease in throughput times diminishes as the fleet size increases. In fact, using dual-load vehicles with an increased fleet size does not lead to a significant increase in performance.

From Table 11 it is clear that using two dual-load vehicles leads to smaller average load throughput times than three uni-load vehicles in the High throughput case (this was the reverse for Random and Structured shifts). Apparently, adding vehicle capacity in an environment with high vehicle utilization has a greater impact on the performance than adding capacity in environments with low vehicle utilization. In this case, the performance still significantly improves when three uni-load vehicles are replaced with three dual-load vehicles. Although the differences in average load throughput times for on-line control are rather small, the phenomenon that FCFS generally leads to smaller average load throughput times in a symmetrical layout compared to a less symmetrical layout, and NVF generally leads to smaller average load throughput times in a less symmetrical layout compared to a symmetrical layout seems to occur for the dual-load vehicles case as well.

Table 11. Average load throughput times for High throughput shifts (60 jobs)

	Off-line control		On-line control: NVF				On-line control: FCFS			
	Insertion: 2 GVs	Insertion: 3 GVs	2 GVs cap. 1	2 GVs cap. 2	3 GVs cap. 1	3 GVs cap. 2	2 GVs cap. 1	2 GVs cap. 2	3 GVs cap. 1	3 GVs cap. 2
U-lay-out										
Average	6422.8	2006.3	6907.3	2012.6	2355.1	1477.5	6805.0	2052.4	2417.8	1470.8
St. dev.	689.0	361.3	700.0	196.4	382.4	96.1	582.0	237.0	412.1	103.2
Deviation	-	-	8 %	-69 %	17 %	-26 %	6 %	-68 %	21 %	-27 %
I-lay-out										
Average	6073.2	2095.2	6518.7	1998.8	2592.3	1328.3	7265.9	2033.5	2586.6	1334.0
St. dev.	730.0	847.8	710.2	278.5	834.1	88.5	757.3	368.4	791.3	109.8
Deviation	-	-	7 %	-67 %	24 %	-37 %	20 %	-67 %	23 %	-36 %

5 Concluding Remarks

In this paper we compared the average load throughput times of several off-line control and on-line dispatching rules for guided vehicles used for internal transport. Using off-line control means that all information on load release times, origins and destinations has to be known in advance. This is not a real situation found in practice due to the stochastic nature of internal transportation environments. However, for theoretical purposes we assumed that all information is available when off-line control rules are used. The performance is defined as the average load throughput time; the time needed to serve a transport request from the moment a load is (physically) released to the system and ready for transport until it is dropped off at its destination, (i.e., the load waiting time plus the load transportation time).

Our goal was to study the performance gap (difference in average load throughput times) between on-line control and off-line dispatching and to investigate how this gap is affected when the fleet size is increased and when the on-line vehicle capacity is increased -in combination with increasing the fleet size, for different dispatching rules in different layout environments.

The results show that for different studied layouts and shifts, considerable gains on performance (reductions in average load throughput times) are possible with off-line algorithms (exact and heuristics) if the system is relatively quiet, i.e. dispatch requests are spread out evenly (low throughput) and vehicles have relatively high idle times (in this case about 15% or more). This is due to reductions in load waiting time by already traveling to a load before it has been physically released.

Therefore the load can be picked up relatively sooner, which leads to a reduction in average load waiting times and in most cases the average load throughput times. In low throughput systems, we see that the fleet size has to increase by 50% or more to obtain similar results to the Optimal routing. However, in systems with high throughput, and therefore a smaller opportunity to reduce load waiting time, the performance of on-line control is already satisfactory (in our case differences of 6-20%).

Table 12 summarizes the results for adding extra vehicle capacity when two or three vehicles are used. (The value between brackets represents the performance of *three* off-line controlled uni-load vehicles relative to *two* off-line controlled uni-load vehicles). It is clear that heavily utilized GVs benefit most from adding vehicle capacity. The benefits decrease as the fleet size increases.

Table 12. Average load throughput time deviations between off-line control (Insertion) and on-line dispatching for changes in fleet capacity

Control form: vehicle types	Performance deviation by heavily utilized GVs (High throughput shifts)	Performance deviation by GVs with idle time (Structured shifts)
U-lay-out	NVF / FCFS	NVF / FCFS
Off-line: 2 uni-load GVs	-	-
On-line: 2 uni-load GVs	8 % / 6 %	26 % / 27 %
On-line: 2 dual-load GVs	-69 % / -68 %	18 % / 18 %
Off-line: 3 uni-load GVs	(-69 %)	(-25 %)
On-line: 3 uni-load GVs	17 % / 21 %	35 % / 31 %
On-line: 3 dual-load GVs	-26 % / -27 %	34 % / 28 %
I-lay-out	NVF / FCFS	NVF / FCFS
Off-line: 2 uni-load GVs	-	-
On-line: 2 uni-load GVs	7 % / 20 %	30 % / 31 %
On-line: 2 dual-load GVs	-67 % / -67 %	22 % / 21 %
Off-line: 3 uni-load GVs	(-65 %)	(-25 %)
On-line: 3 uni-load GVs	24 % / 23 %	32 % / 32 %
On-line: 3 dual-load GVs	-37 % / -36 %	31 % / 32 %

In Table 13, we can also see the effects on on-line performance when the fleet capacity increases, compared with two GVs controlled off-line. In the case of high throughput environments, the difference between off-line control and on-line dispatching performance is in the standard situation already almost negligible (8% or less, as can be seen the second column of Table 13). In fact, adding vehicles to the fleet or adding capacity to the vehicles improves the performance beyond the off-line (standard) situation.

Table 13. Average load throughput time deviations for several situations of on-line dispatching relative to off-line control (Insertion) with two uni-load vehicles

Situation	Performance deviation by heavily utilized GVs (High throughput shifts)	Performance deviation by GVs with idle time (Structured shifts)
U-lay-out	NVF / FCFS	NVF / FCFS
Standard (2 GVs)	8 % / 6 %	26 % / 27 %
50 % extra GVs	-63 % / -62 %	2 % / -2 %
Dual-load GVs	-69 % / -68 %	18 % / 18 %
50 % extra + dual-load GVs	-77 % / -77 %	0.3 % / -4 %
I-lay-out	NVF / FCFS	NVF / FCFS
Standard (2 GVs)	7 % / 20 %	30 % / 31 %
50 % extra GVs	-57 % / -57 %	-2 % / -2 %
Dual-load GVs	-67 % / -67 %	22 % / 21 %
50 % extra + dual-load GVs	-78 % / -78 %	-2 % / -2 %

- Increasing the fleet size by 50% for the Structured shifts with vehicle idle time (relatively low throughput environment), leads to a performance deviation of 2% for NVF and -2% for FCFS from the off-line heuristic. However, this is still better than doubling the vehicle capacity, which leads to a performance deviation of 18%. The effect of increasing the number of vehicles in a low throughput environment is so dominant that combining the effects of extra vehicles plus extra vehicle capacity does not lead to a combined increase in performance. Table 13 reveals which steps could be taken to close the performance gap between off-line control and on-line dispatching for a certain environment.
- Careful study also reveals that the NVF rule seems to perform more often more favorable in a non-symmetric layout environment and FCFS more favorable in a symmetric environment. This seems logical since decisions based on symmetric distances with the NVF rule are similar to random load-to-vehicle assignments. Furthermore, the standard deviations of the average load throughput times are in general higher for on-line control compared to off-line control and decreases as the fleet capacity increases. This is due to the phenomenon that when the fleet capacity is relatively greater during peak periods of load releases, the maximum load waiting times decrease.

References

Ball, M.O. / Magnanti, T.L. / Monma, C.L. / Nemhauser, G.L., (1995): Network Routing. In: *Handbooks in operations research and management science*, Elsevier, Vol. 8.

Bozer, Y A. / Cho, M. / Srinivasan, M M. (1994): Expected waiting times in single-device trip-based material handling systems. In: *European Journal of Operational Research*, Vol. 75, p. 200-216.

De Koster, R. / Van der Meer, J.R. (1998): Centralized versus Decentralized Control of Internal Transport, a case study. In: B. Fleischmann / J.A.E.E. van Nunen / M. Grazia

Speranza and P. Stähly (Eds.), *Advances in distribution logistics*, pp. 403-420. Springer, Berlin.

Dumas, Y. / Desrosiers, J. / Soumis, F. (1991): The pickup and delivery problem with time windows. In: *European Journal of Operations Research,* Vol. 54, p. 7-22.

Gendreau, M. / Hertz, A. / Laporte, G. (1992): New insertion and postoptimization procedures for the traveling salesman problem. In: *Operations Research*, Vol. 40, No. 6, pp. 1086-1094.

Jaw, J. / Odini, A. / Psaraftis, H. / Wilson, N. (1986): A heuristic algorithm for the multi-vehicle advance-request dail-a-ride problem with time windows. In: *Transportation Research B*, Vol. 20, No. 3, pp. 243-246.

Miller, C. / Tucker, A. / Zemlin, R. (1960): Integer programming formulations and traveling salesman problems. In: *Journal of the ACM*, Vol. 7, p. 326-329.

Savelsbergh, M.W.P. / Sol, M. (1995): The general pickup and delivery problem. In: *Transportation Science*, Vol. 29, No. 1, p. 17-29.

Solomon, M. (1987): The vehicle routing and scheduling problem with time window constraints. In: *Operations Research*, Vol. 35, p. 254-265.

Solomon, M.M. / Desrosiers, J. (1988): Time window constrained routing and scheduling problems. In: *Transportation Science*, Vol. 22, No. 1, p. 1-13.

Srinivasan, M.M. / Bozer, Y.A. / Cho, M. (1994): Trip-based material handling systems: throughput capacity analysis. In: *IIE Transactions*, Vol. 26, No. 1, p. 70-89.

Tompkins, J.A. / White, J.A. / Bozer, Y.A. / Frazelle, E.H. / Tanchoco, J.M.A. / Trevino, J. (1996): Facilities Planning. 2nd ed., John Wiley & Sons, Inc., New York.

Van der Meer, J.R. / De Koster, R. (1998): A Classification of Control Systems for Internal Transport. In: *Progress in Material Handling Research*, Graves *et al.* (eds.) Ann Arbor, Michigan, p. 633-650. ISBN 1-882780-03-5.

Average Costs versus Net Present Value: A Comparison for Multi-source Inventory Models

Erwin van der Laan[1] and Ruud Teunter[2]

[1] Rotterdam School of Management, Erasmus University Rotterdam
P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands
[2] Econometric Institute, Erasmus University Rotterdam
P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands

Abstract While the net present value (NPV) approach is widely accepted as the right framework for studying production and inventory control systems, average cost (AC) models are more widely used. For the well known EOQ model it can be verified that (under certain conditions) the AC approach gives near optimal results, but does this also hold for more complex systems? In this paper it is argued that for more complex systems, like multi-source systems, one has to be extremely careful in applying the AC approach on intuition alone, even when these systems are deterministic. Special attention is given to a two-source inventory system with manufacturing, remanufacturing, and disposal, and it is shown that for this type of models there is a considerable gap between the AC approach and the NPV aprroach.

Keywords: Net present value, average costs, inventory control, manufacturing, remanufacturing, disposal, holding costs.

1 Introduction

Several authors (e.g. Hadley, 1964; Trippi and Lewin, 1974; Thompson, 1975; Hofmann, 1998; Klein Haneveld and Teunter, 1998) have argued that *for the EOQ model* the average cost (AC) framework as an approximation to the superior net present value (NPV) framework leads to near optimal results under the following conditions:

- Products are not moving too slow,
- Interest rates are not too high,
- The customer payment structure does not depend on the inventory policy.

The first two conditions have to guarantee that compounded interest does not effect the results too much. That the latter condition is crucial was first put forward by Beranek (1966), who's concern was confirmed later by Grubbström (1980) and Kim *et al.* (1984).

The main objections against the average cost approach, as it is usually applied as an approximation to the net present value approach, are threefold:

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

O1 The time value of money is not explicitly taken into account,
O2 There is no distinction between out-of-pocket holding costs and opportunity costs due to inventory investment, while other sources of opportunity costs/yields (fixed ordering costs, product sales) are not taken into account at all.
O3 Initial conditions are not taken into account

Yet, the net present value approach is often rather complicated, so an approximation may still be preferred.

Several authors have tried to deal with the above problems by showing that a certain transformation of the holding cost parameters in EOQ-type models gives near optimal results from an NPV perspective. This, however, shifts the problem to finding the right transformation. Up to now only ad hoc solutions have been given that are often very counter-intuitive (see e.g. Beranek, 1966; Corbey *et al.*, 1999; Luciano and Peccati, 1999). No general principle has been developed to solve the transformation problem.

This paper intends to systematically analyze the differences between the AC and NPV approach and its consequences for modeling inventory systems. To that end we will analyze a number of deterministic models with increasing complexity, starting with the standard EOQ model and moving towards multi-echelon and multi-source models. It is shown that there are basically two classes of systems: 1. systems for which a transformation of model parameters exists that is independent of decision variables, such that the AC approach and the NPV approach are approximately equal, and 2. systems for which such a transformation does not exist.

This paper is further organized as follows: In the next section we propose a general principle that allows us to handle the NPV approach for deterministic systems in a very simple way. Moreover, with this principle we can easily compare the AC approach with the NPV approach. We then show how the NPV approach compares to the AC approach for the EOQ model (Section 3), multi-echelon systems (Section 4), and multi-source systems (Section 5). The theoretical results are further illustrated by a small numerical study in Section 6. We end with a summary and discussion of the main results in Section 7.

2 A General Principle for the NPV Approach in Deterministic Models

We define the *Net Present Value* (NPV) as the total discounted cash-flow over an infinite horizon. Additional to the NPV we define the *Annuity Stream* (AS) as

$$AS = r\{NPV\}.$$

where r denotes the discount rate. The annuity stream is the transformation of a set of discrete and/or continuous cash flows to one continuous stream of cash-flows, such that the latter has the same net present value as the original set of cash-flows. The notion of an annuity stream is useful, since it can be directly compared with average costs.

If T denotes the cycling time of a discrete cash-flow C, with first occurrence time T_1, then the annuity stream is given by

$$AS = rC \sum_{n=0}^{\infty} \mathrm{e}^{-r(T_1+nT)} = \frac{rC\mathrm{e}^{-rT_1}}{1-\mathrm{e}^{-rT}}, \tag{1}$$

This can be written as the McLaurin expansion

$$AS = \frac{rC\mathrm{e}^{-rT_1}}{1-\mathrm{e}^{-rT}} = \frac{C}{T} + C\left[r\left(\frac{1}{2} - \frac{T_1}{T}\right) + O(r^2 \max\{T, T_1\})\right],$$

so that we have the following linearisation in r of the annuity stream:

$$\overline{AS} = \frac{C}{T} + rC\left(\frac{1}{2} - \frac{T_1}{T}\right). \tag{2}$$

Note that in most practical applications r is small and $0 \leq T_1 \leq T$, so that the above approximation is quite reasonable.

The first term of (2), C/T, denotes the average cash-flow per time unit, as it would follow from a standard AC calculation. The second term may be viewed as a first order correction term to account for the time value of money. This is graphically shown in 1. Approximately, the AC approach underestimates the interest component of the annuity stream if $T_1 \leq T/2$ and overestimates otherwise. The results of both approaches are the same if $T_1 \approx T/2$.

The above only holds for discrete cash-flows, but we can do a similar analysis for continuous cash-flows. Suppose that a product is sold with continuous rate λ for a price p, starting at time T_1. Then, its annuity stream is given by

$$\begin{aligned} AS &= rp\lambda \textstyle\int_{T_1}^{\infty} \mathrm{e}^{-rt}\mathrm{d}t \\ &= rp\lambda \mathrm{e}^{-rT_1} \\ &= p\lambda[1 - rT_1 + O(r^2{T_1}^2)] \\ &\approx p\lambda[1 - rT_1]. \end{aligned} \tag{3}$$

The way that the AC approach usually deals with the *underestimation* of the interest component for cash-flows related to variable production costs is to add a certain factor to the out-of-pocket holding cost parameter. This

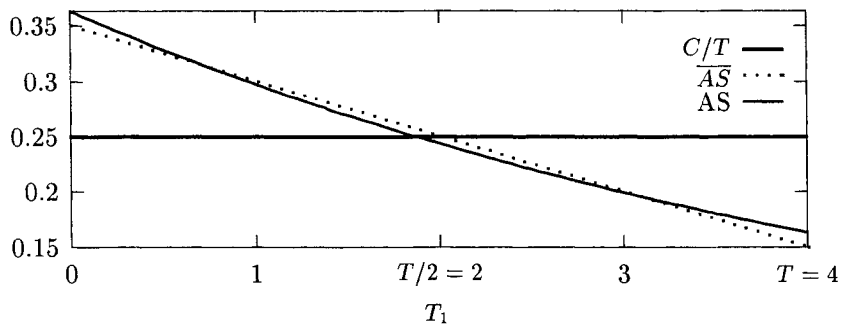


Fig. 1. Comparison between the average cash-flow per time unit (C/T), the annuity stream (AS), and its linearisation ($\overline{AS}$) for $T = 4$, $C = 1$, and $r = 0.2$

factor is usually taken as the interest rate r times the 'value' of the stocked item. This approach has a number of disadvantages. First, it assumes that the overestimation is proportional with average inventory. We will show that this does not need to be the case. In fact, size and timing of cash-flows are dependent on *cycle times* rather than the existence of physical stocks. Second, it only deals with *under*estimation of the interest component and not with *over*estimation, since the value of a stocked item is usually taken to be positive. Third, this approach only considers the interest components of variable production costs, while interest components of all other cash flows (fixed costs, sales, etc.) are not taken into account. Finally, it is unclear what is meant by the 'value' of a stocked item, since this depends on the type of decision that has to be made.[1]

3 From NPV to AC with the EOQ Model

First consider the basic EOQ model in an NPV framework (Figure 2). Demand for a product with selling price p is continuous with rate λ, generating a continuous cash inflow of λp per time unit. Every T periods a batch of Q products is produced against variable cost c per product and fixed cost K per batch (zero lead time) starting at time $t = 0$. To keep the analysis simple and transparent we will not consider out-of-pocket holding costs. Note that in the AC framework holding costs appear as an approximation to the annuity

[1] Depending on the type of decision that has to be made one could say that a product return has value zero if it has been obtained for free. At the same time one could say it has value $c_p - c_r$, since after remanufacturing against cost c_r it can be sold for c_p. As a third option one could say that its value is $c_m - c_r$, since this is the difference between manufacturing against cost c_m and remanufacturing against cost c_r.

stream to account for interest components. We will refer to this holding cost parameter as the 'opportunity cost rate of inventory investment'.

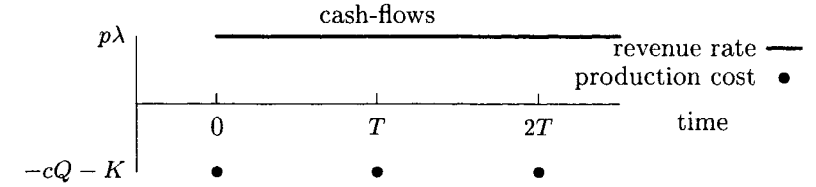


Fig. 2. Relevant cash-flows for the EOQ model with continuous demand

The total annuity stream for this deterministic system consists of the annuity stream due to *a)* the variable revenues and production costs (AS_v)

$$\begin{aligned} AS_v &= r\left(p\lambda \int_0^{\infty} \mathrm{e}^{-rt}\mathrm{d}t - cQ\sum_{n=0}^{\infty}\mathrm{e}^{-rnT}\right) \\ &= p\lambda - \frac{rcQ}{1-\mathrm{e}^{-rT}} \\ &\approx (p-c)\lambda - rcQ/2, \end{aligned} \tag{4}$$

and *b)* the annuity stream due to fixed set-up costs (AS_f)

$$\begin{aligned} AS_f &= -r\textstyle\sum_{n=0}^{\infty} K\mathrm{e}^{-rnT} \\ &= -\tfrac{rK}{1-\mathrm{e}^{-rT}} \\ &\approx -K\lambda/Q - rK/2, \end{aligned} \tag{5}$$

where we have used linearizations (2) and (3) with $T_1 = 0$. Combining (4)–(5) we arrive at the approximated total annuity stream function

$$\overline{AS} = (p-c)\lambda - K\lambda/Q - rcQ/2 - rK/2 \tag{6}$$

The first term in (6) denotes marginal net profits per time unit, and the second term denotes the average set-up costs per time unit. The other terms are interest components.

The standard AC approach calculates the average profit (AP) function as

$$AP = (p-c)\lambda - hQ/2 - K\lambda/Q, \tag{7}$$

where h is the holding cost rate to account for the opportunity costs of inventory investment. Optimizing (7) leads to the well-known EOQ formula, but it is not immediately clear what the value of h should be. However, if we want that optimizing AP gives the same order size as optimizing $\overline{AS}$ we should choose $h = rc$. Although this value will appeal to most people's intuition it is important to note that more complicated models, as the ones encountered in the remainder of the paper, call for more complicated holding cost rates for which an intuitive explanation is often hard to give.

4 Multi-Echelon Systems

Consider a two-echelon system (Figure 3) consisting of processes i, $i \in \{1, 2\}$, with fixed processing time L_i per batch, unit processing cost c_i and fixed processing cost K_i. A production batch of size Q is initiated every T time units starting at time $T_1 = 0$. As soon as process 1 finishes process 2 starts. As soon as process 2 finishes, the batch enters serviceable inventory. The relevant cash flows are depicted in Figure 4. Note that production costs are incurred at the beginning of each process and that product sales only start after the first production batch has entered the serviceable inventory, i.e., at time $L_1 + L_2$.

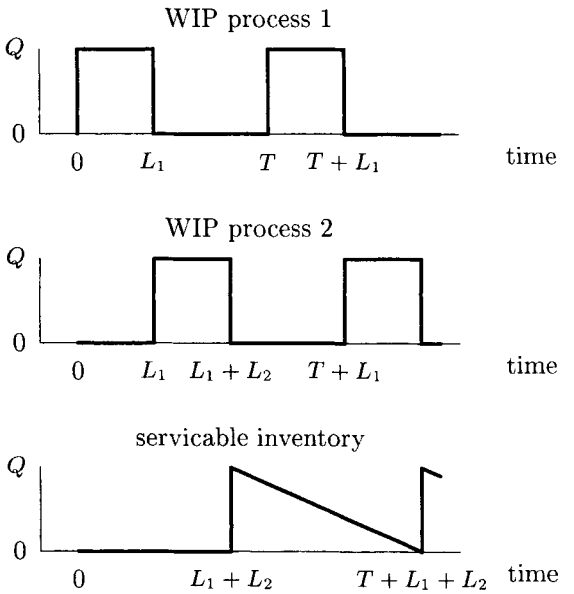


Fig. 3. The inventory processes of a two-echelon system

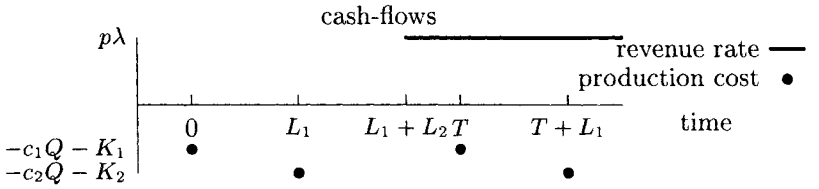


Fig. 4. Relevant cash-flows for the two-echelon system

A formal deduction of AS_v gives

$$AS_v = r\left(\frac{p\lambda e^{-r(L_1+L_2)}}{r} - \frac{Qc_1 + Qc_2e^{-rL_1}}{1-e^{-rT}}\right)$$

$$\approx (p - c_1 - c_2)\lambda - r\left((c_1 + c_2)Q/2 - c_2L_1\lambda\right) - rp(L_1 + L_2)\lambda. \tag{8}$$

The first term in (8) is just the marginal net profits per time unit, whereas the second term denotes the opportunity costs of inventory investment. The last term represents the opportunity costs of delayed product sales.

The traditional average cost approach would calculate the average profit function as the average net marginal profits per time unit minus the average holding costs per time unit,

$$AP_v = (p - c_1 - c_2)\lambda - h_1L_1\lambda - h_2L_2\lambda - h_sQ/2, \tag{9}$$

where the second term is the average work in process inventory of process 1 charged with opportunity holding cost rate h_1, the third term is the average work in process inventory of process 2, charged with rate h_2, and the fourth term is the average serviceable inventory charged with rate h_s. Equation (9) corresponds to (8) if we employ the following transformation of cost parameters:

$$h_1 \to r(p - c_2)$$

$$h_2 \to rp$$

$$h_s \to r(c_1 + c_2)$$

The parameter h_s can be interpreted intuitively as the interest rate times the total marginal production costs. The other holding cost rates are less intuitive, but that is not really a problem since for any value of these parameters the difference between NPV and AC will merely be a constant.

The annuity stream due to fixed set-up costs is

$$AS_f = -r\left(\frac{K_1 + K_2 e^{-rL_1}}{1 - e^{-rT}}\right)$$
$$\approx -\frac{(K_1 + K_2)\lambda}{Q} - r\left(\frac{K_1 + K_2}{2} - \frac{K_2 L_1 \lambda}{Q}\right).$$

In the traditional average cost approach opportunity costs of set-ups are never explicitly taken into account (compare to the EOQ model, where opportunity costs of set-ups are a constant and can be left out). Here, however, we see that the opportunity costs do depend on the order size Q and can no longer be discarded. Again, we can map (up to a constant) the average cost approach to the linearization of the annuity stream by the transformation

$$K_1 \to K_1$$
$$K_2 \to K_2(1 - rL_1)$$

Summarizing, we can say that the traditional average cost approach is still applicable for multi-echelon structures, as long as the right transformations of model parameters are used. These transformations, however, depend on initial conditions and are not very intuitive. Thus, to find the correct transformations we have to rely on an NPV analysis. It is comforting though that for this class of models the traditional average cost models can still be applied.

5 Multi-Source Systems

Until now we only considered situations in which inventories consist of products that all have generated the same cash-flows. Additional problems may arise if inventories consist of products that have been produced in different ways against different costs. This is the case with products that can be both newly manufactured and remanufactured from old products. Remanufactured products have the same functionality and quality as newly produced products and can therefore be sold at the same market for the same price. In this sense they are indistinguishable and can be put in the same inventory. However, the cash flows generated by manufactured products are different from remanufactured products, since they follow from different processes with different costs. In this section we show how this affects the difference between NPV and AC.

5.1 A System with Manufacturing and Remanufacturing

Consider a two source system (Figure 5), where product demand can be fulfilled both by manufactured products, with marginal cost c_m and fixed set-up cost K_m, and remanufactured products, with marginal cost c_r and

fixed set-up cost K_r. Manufactured and remanufactured products have the same quality standards and are sold on the same market against the same price p with rate λ. The main difference between the manufacturing process and remanufacturing process is that the latter depends on the flow of product returns, which for now is assumed to be deterministic with rate γ, $0 < \gamma < \lambda$.

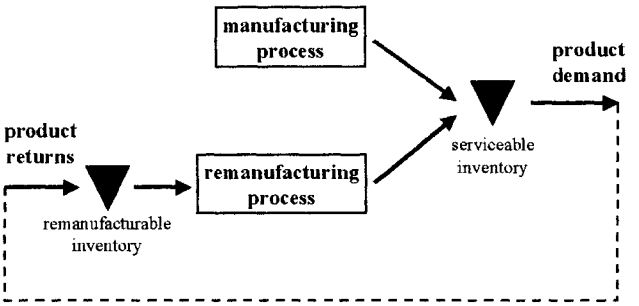


Fig. 5. Schematic representation of a manufacturing/remanufacturing system

This system was first proposed by Schrady (1967) and further analyzed by Richter (1996) and Teunter (1998). In the above-mentioned papers the system is controlled by subsequently producing N manufacturing batches and M remanufacturing batches. For ease of explanation we assume here that $N = M = 1$ so that the system is controlled by repeatedly producing one manufacturing batch of size Q_m, succeeded by one remanufacturing batch of size Q_r.[2] We assume that at time 0 we start with zero inventory of both serviceables and remanufacturables. Thus, to start up the system and to guarantee a monotonous ordering strategy at the same time, we have to start with a *manufacturing* batch of size Q_r. The first *regular* manufacturing batch of size Q_m then occurs at time $T_r = Q_r/\lambda$ and the first remanufacturing batch occurs at time $T = (Q_m + Q_r)/\lambda$. Continuing this way, manufacturing batches and remanufacturing batches occur every T time units. Leadtimes are assumed to be zero. Note that, since all returns are used for remanufacturing, we have $Q_r = \gamma T$, $Q_m = (\lambda - \gamma)T$ and $Q_r = \frac{\gamma}{\lambda-\gamma} Q_m$. The timing of all relevant cash-flows is visualized in Figure 6.

The AS_v for this system reads

$$AS_v = p\lambda - r\left(Q_r c_m + \frac{Q_m c_m \mathrm{e}^{-rT_r} + Q_r c_r \mathrm{e}^{-rT}}{1 - \mathrm{e}^{-rT}}\right)$$

$$\approx (p\lambda - c_m(\lambda - \gamma) - c_r\gamma) - rc_m Q_m\left(\tfrac{1}{2} - \tfrac{\gamma}{\lambda}\right) - r\left(c_m - c_r/2\right)Q_r. \quad (10)$$

[2] The following analysis is easily extended to arbitrary N and M, but this would only lengthen the mathematical expressions without gaining additional insight.

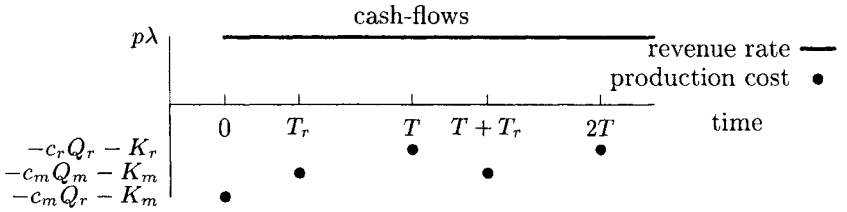


Fig. 6. Relevant cash-flows for the two-source system without disposal

Again, the first term denotes the total marginal profits and the last two terms denote the total opportunity costs of inventory investment.

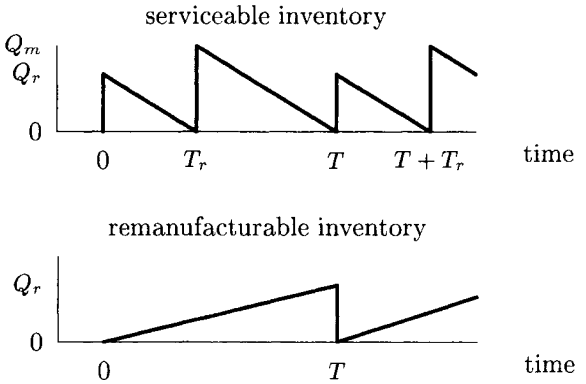


Fig. 7. The inventory processes of a two-source system

Let's now compare the above expression with the corresponding average profit function. Figure 7 depicts the inventory process of serviceables and remanufacturables, from which we derive that the long run average inventory of remanufactured products equals $Q_rT_r/(2T) = (\gamma/\lambda)Q_r/2$, the long run average inventory of manufactured products equals $Q_m(T-T_r)/(2T) = (1-\gamma/\lambda)Q_m/2$, and the average inventory of remanufacturable products equals $Q_r/2$. This leads to the following average profit function:

$$AP_v = (p\lambda - c_m(\lambda-\gamma) - c_r\gamma) - h_m\left(1-\tfrac{\gamma}{\lambda}\right)Q_m/2 - h_r\left(\tfrac{\gamma}{\lambda}\right)Q_r/2 - h_nQ_r/2. \tag{11}$$

Clearly, both opportunity costs and average inventory are linear in Q_m and Q_r so that h_m, h_r, and h_n can be chosen such that (11) is equivalent to (10). Since the average inventory of remanufactured products and the average inventory of remanufacturables are both linear in Q_r, either h_r or h_n is redundant. Naturally we choose $h_n = 0$ because there are no investments in remanufacturable inventory and thus no associated opportunity cost exist. This gives

$$\begin{aligned} h_m &\to rc_m \left(\tfrac{\lambda}{\lambda-\gamma}\right) \\ h_r &\to rc_r \left(2 - \tfrac{\lambda}{\gamma}\right) \\ h_n &\to 0 \,. \end{aligned}$$

So, setting $h_n = 0$ leads to different holding cost rates for manufactured and remanufactured products. This is rather counter-intuitive and may lead to the (false) conclusion that in fulfilling product demands priority should be given to either manufactured or remanufactured products, whichever generates more opportunity costs. That this conclusion is false can be clearly seen when we look at it from an NPV perspective. The financial consequences of selling either a manufactured item or a remanufactured item are exactly the same, since they generate the *same cash inflow* at the *same time.*

These counter-intuitive results can be avoided by choosing the same holding cost rates for manufactured and remanufactured products. This gives

$$\begin{aligned} h_m &\to rc_m \\ h_r &\to rc_m \\ h_n &\to r(c_m - c_r) \,. \end{aligned}$$

What about the set-up costs? The AS_f is derived as

$$\begin{aligned} AS_f &= -r\left(K_m + \frac{K_m \mathrm{e}^{-rT_r} + K_r \mathrm{e}^{-rT}}{1 - \mathrm{e}^{-rT}}\right) \\ &\approx -(\lambda-\gamma)K_m/Q_m - \gamma K_r/Q_r - rK_m\left(\tfrac{3}{2} - \tfrac{\gamma}{\lambda}\right) + rK_r/2, \end{aligned} \tag{12}$$

and we observe that the opportunity costs of set-ups do not depend on the policy parameters. This however will change in the next section.

Our framework shows that the only way to influence the opportunity costs of holding inventories is to somehow change the *timing* of the investments c_m and c_r, for instance by using pull and push type policies (see van der Laan *et al.*, 1998), or to somehow change the *fraction* of (re)manufactured products by using a disposal policy (see e.g. Inderfurth, 1997; van der Laan and Salomon, 1997). In the next paragraph we extend the manufacturing/remanufacturing model with the option to dispose product returns.

5.2 A System with Manufacturing, Remanufacturing, and Disposal

A number of authors have considered disposal strategies in a manufacturing/remanufacturing environment in order to optimize total system costs (e.g. Heyman, 1977; Inderfurth, 1997; Richter,1996; Simpson, 1978; van der Laan and Salomon, 1997). However, doing so, care should be taken in the modeling process.

Consider the example of Section 5.1, but instead of remanufacturing all product returns we decide to use only a fraction U, $0 \leq U \leq 1$, and continuously dispose a fraction $1 - U$. The unit 'cost' related to disposal, c_d can be positive (for instance if products contain hazardous materials, which need to be processed in an environmental friendly manner), or negative (for instance if product returns have a positive salvage value and can be sold to a third party). Define the decision variable $\Gamma = U\gamma$, then the total amount of disposals during a production cycle of length T equals $(\gamma - \Gamma)T$.

The AS_v for the situation with continuous disposals is

$$\begin{aligned} AS_v &= p\lambda - (\gamma - \Gamma)c_d - r\left(Q_r c_m + \tfrac{Q_m c_m e^{-rT_r} + Q_r c_r e^{-rT}}{1 - e^{-rT}}\right) \\ &\approx (p\lambda - c_m(\lambda - \Gamma) - c_r\Gamma - c_d(\gamma - \Gamma)) \\ &\quad -rc_m Q_m\left(\tfrac{1}{2} - \tfrac{\Gamma}{\lambda}\right) - r\left(c_m - c_r/2\right)Q_r\,. \end{aligned}$$

The parameter c_d only appears in the marginal cost term so there are no opportunity costs associated with product disposal.

If $c_d > 0$, it is more efficient from a financial point of view to dispose as late as possible. One could choose to dispose a batch of remanufacturables whenever a certain capacity limit has been reached. Here, we choose to accumulate the products to be disposed and dispose them all at once whenever a remanufacturing batch is initiated, i.e. at time T_m, $T + T_m$, and so on. The amount disposed at the end of each cycle equals $(\gamma - \Gamma)T$. The AS_v for batch-disposals thus reads

$$\begin{aligned} AS_v &= p\lambda - r\left(Q_r c_m + \tfrac{Q_m c_m e^{-rT_r} + (Q_r c_r + (\gamma - \Gamma)Tc_d)e^{-rT}}{1 - e^{-rT}}\right) \\ &\approx (p\lambda - c_m(\lambda - \Gamma) - c_r\Gamma - c_d(\gamma - \Gamma)) \\ &\quad -rc_m Q_m\left(\tfrac{1}{2} - \tfrac{\Gamma}{\lambda}\right) - r\left(c_m - c_r/2\right)Q_r + rc_d(Q_m + Q_r)\left(\tfrac{\gamma - \Gamma}{2\lambda}\right)\,. \end{aligned} \tag{13}$$

Note the opportunity cost/yield related to disposal, $rc_d(Q_m + Q_r)\left(\frac{\gamma - \Gamma}{2\lambda}\right)$.

The annuity stream due to (re)manufacturing set-ups is derived as

$$AS_f = -r\left(K_m + \frac{K_m e^{-rT_r} + K_r e^{-rT}}{1 - e^{-rT}}\right)$$

$$\approx -(\lambda - \Gamma)K_m/Q_m - \Gamma K_r/Q_r - rK_m\left(\tfrac{3}{2} - \tfrac{\Gamma}{\lambda}\right) + rK_r/2. \tag{14}$$

We observe that the opportunity costs of set-ups depend on the policy parameter Γ, and can no longer be ignored (compare with (12)).

Combining (13) and (14) we find that for $0 < \Gamma < \lambda$ the total annuity stream is given by

$$\begin{aligned} AS = {} & p\lambda - r(K_m + Q_r c_m) \\ & - r\left(\frac{(K_m + Q_m c_m)e^{-rT_r} + (K_r + Q_r c_r + (\gamma - \Gamma)Tc_d)e^{-rT}}{1 - e^{-rT}}\right), \end{aligned} \tag{15}$$

which can be approximated by the function

$$\begin{aligned} \overline{AS} = {} & p\lambda - c_m(\lambda - \Gamma) - c_r\Gamma - c_d(\gamma - \Gamma) \\ & -(\lambda - \Gamma)K_m/Q_m - \Gamma K_r/Q_r \\ & -rK_m - r(K_m + c_m Q_m)\left(\tfrac{1}{2} - \tfrac{\Gamma}{\lambda}\right) + rK_r/2 - r(c_m - c_r/2)Q_r \\ & +rc_d(Q_m + Q_r)\left(\tfrac{\gamma - \Gamma}{2\lambda}\right). \end{aligned} \tag{16}$$

The traditional AC approach calculates the total average profit function as the total marginal profits, set-up costs, and inventory costs as

$$\begin{aligned} AP = {} & p\lambda - c_m(\lambda - \Gamma) - c_r\Gamma - c_d(\gamma - \Gamma) \\ & -(\lambda - \Gamma)K_m/Q_m - \Gamma K_r/Q_r \\ & -h_m(1 - \tfrac{\Gamma}{\lambda})Q_m/2 - h_r\left(\tfrac{\Gamma}{\lambda}\right)Q_r/2 - h_n\left(\tfrac{\gamma}{\lambda}\right)(Q_m + Q_r)/2. \end{aligned} \tag{17}$$

Using the relation $Q_r = \frac{\Gamma}{\lambda - \Gamma}Q_m$ it is easily verified that we can transform AP into $\overline{AS}$ (up to a constant), by using the following transformations of c_r, h_m, h_r, and h_n:

$$\begin{aligned} c_m &\to c_m + rK_m/\lambda \\ h_m &\to rc_m \\ h_r &\to rc_m \\ h_n &\to r\left[(\Gamma/\gamma)(c_m - c_r) - (1 - \Gamma/\gamma)c_d\right] \end{aligned} \tag{18}$$

Clearly, this is a non-linear transformation in the decision variable Γ, which indicates the considerable gap between the traditional average cost approach and the linearization of the annuity stream. This is further illustrated by some analytical and numerical results in the next section.

6 Analytical and Numerical Comparison of Alternative Transformations

Consider the inventory system with manufacturing, remanufacturing, and batch-disposal of Section 5.2. In this section we investigate how the average cost approach performs with respect to the linearization of the annuity stream approach, when forcing a linear transformation of the cost parameters that does not depend on decision variables.

Since demand is either fulfilled by manufacturing or remanufacturing and the number of (re)manufacturing batches within production cycle T is fixed to one, we have $Q_r = \frac{\Gamma}{\lambda - \Gamma} Q_m$. Hence, for $0 < \Gamma \leq \gamma < \lambda$ expressions (15) – (17) can be transformed into functions of Q_m and Γ only. For the special case $\Gamma = 0$ these functions are derived in the appendix. For all the numerical examples in this section we use the base-case scenario of Table 1, unless specified otherwise.

Table 1. Base case scenario

parameter	λ	γ	p	c_m	c_r	c_d	K_m	K_r	r
value	20	10	20	10	5	5	10	10	0.10

As a performance measure for batch size Q we define the relative difference

$$R(Q) = \left[1 - \frac{\tilde{AS}(Q)}{\tilde{AS}(Q^{\mathrm{AS}})}\right] \times 100\%,$$

where $\tilde{AS}(.) = AS(.) - (p\lambda - c_m(\lambda - \gamma) - c_r\gamma)$ is the relevant annuity stream and Q^{AS} is the batch-size that maximizes $AS(.)$.

In our analysis we consider two transformations.

Transformation A An intuitive, though rather naive, transformation is the following:

$$h_m \rightarrow rc_m$$

$$h_r \rightarrow rc_r$$

$$h_n \rightarrow 0$$

The above choice follows from the (false) intuition that opportunity costs of inventory investment are (approximately) equal to the interest rate times the average inventory investment. Parameter c_r is chosen according to (18)) to take the opportunity cost of remanufacturing batches into account:

$$c_m \to \begin{cases} c_m + rK_m/\lambda, & \text{if } \Gamma > 0 \\ c_m, & \text{otherwise} \end{cases} \tag{19}$$

Transformation B A seemingly more sophisticated transformation of h_m, h_r, and h_n was proposed by Inderfurth and Teunter (1998) on the basis of a heuristic argument: "The money tied up in a non-serviceable item is $-cd$, since that could have been 'earned' by disposing of it. Hence, $h_n = r(-c_d)$ (...). The money tied up in a remanufactured item is that tied up in a non-serviceable item plus the cost c_r of remanufacturing the item. Hence, $h_r = r(c_r - c_d)$ (...). The money tied up in a manufactured item is simply the cost c_m of manufacturing an item. Hence, $h_m = rc_m$." Summarizing:

$$\begin{aligned} h_m &\to rc_m \\ h_r &\to r(c_r - c_d) \\ h_n &\to -rc_d \end{aligned}$$

Parameter c_r is chosen according to (19) to take the opportunity cost of remanufacturing batches into account. Note that for $c_d = 0$ *Transformation A* and *Transformation B* are equivalent.

To compare the various approaches, we consider two cases.

Case 1: $\Gamma = 0$ If $\Gamma = 0$ the difference between $\overline{AS}$ and AP is given by

$$\overline{AS} - AP = \left[h_m - rc_m + (h_n + rc_d)\left(\frac{\gamma}{\lambda}\right)\right] Q_m/2 - rK_m/2 \tag{20}$$

where the last term is just a constant but the first term depends on Q_m. If *Transformation B* is applied the first term vanishes, hence the two approaches are equal up to a constant. For *Transformation A* however the righthand side of (20) equals

$$rc_d\left(\frac{\gamma}{\lambda}\right) Q_m/2 - rK_m/2$$

Thus under *Transformation A* the two approaches will differ significantly for large enough $|c_d|$ (see Table 2).

Table 2. Performance of $Q_m^{\overline{AS}}$ and Q_m^{AP} under the base-case scenario for $\Gamma = 0$ and various values of c_d.

					Transform. A		Transform. B	
c_d	Q_m^{AS}	$\tilde{AS}(Q_m^{AS})$	$Q_m^{\overline{AS}}$	$R(Q_m^{\overline{AS}})$	Q_m^{AP}	$R(Q_m^{AP})$	Q_m^{AP}	$R(Q_m^{AP})$
-15	15.1	73.01	15.1	0.0	20.0	1.5	15.1	0.0
-10	16.3	24.94	16.3	0.0	20.0	2.1	16.3	0.0
-5	17.7	-22.97	17.9	0.0	20.0	-0.7	17.9	0.0
0	19.7	-70.67	20.0	0.0	20.0	0.0	20.0	0.0
5	22.4	-118.10	23.1	0.0	20.0	-0.1	23.1	0.0
10	26.6	-165.12	28.3	0.0	20.0	-0.4	28.3	0.0
15	33.8	-211.49	40.0	-0.1	20.0	-0.9	40.0	-0.1

Case 2: $\Gamma = \gamma$ If $\Gamma = \gamma$ the difference between $\overline{AS}$ and AP is given by

$$\begin{aligned}\overline{AS} - AP &= \left[h_m\left(1 - \tfrac{\gamma}{\lambda}\right) - rc_m\left(\tfrac{1}{2} - \tfrac{\gamma}{\lambda}\right)\right] Q_m/2 \\ &+ \left[h_r\left(\tfrac{\gamma}{\lambda}\right) - r(2c_m - c_r) + h_n\right] Q_r/2 - rK_m\left(\tfrac{3}{2} - \tfrac{\gamma}{\lambda}\right) + rK_r/2\,.\end{aligned} \tag{21}$$

Under *Transformation A* the righthand side of (21) reduces to

$$r(c_r - c_m)\left(\frac{\gamma}{\lambda}\right)\left(\frac{\lambda+\gamma}{\lambda-\gamma}\right) Q_m/2 - rK_m\left(\frac{3}{2} - \frac{\gamma}{\lambda}\right) + rK_r/2.$$

The difference will be significant for large enough $|c_m - c_r|$ and/or γ (see Table 3 and 4).

Table 3. Performance of $Q_m^{\overline{AS}}$ and Q_m^{AP} under the base-case scenario for $\Gamma = \gamma$ and various values of c_r. a) The objective function is an increasing function in Q_m

					Transform. A		Transform. B	
c_r	Q_m^{AS}	$\tilde{AS}(Q_m^{AS})$	$Q_m^{\overline{AS}}$	$R(Q_m^{\overline{AS}})$	Q_m^{AP}	$R(Q_m^{AP})$	Q_m^{AP}	$R(Q_m^{AP})$
0	14.2	-28.71	14.1	0.0	28.3	-23.8	$\infty^{a)}$	$-\infty$
5	16.3	-25.00	16.3	0.0	23.1	-6.0	$\infty^{a)}$	$-\infty$
10	19.7	-20.67	20.0	0.0	20.0	0.0	40.0	-26.7
15	26.1	-15.27	28.3	-0.3	17.9	-7.5	28.3	-0.3
20	43.1	- 7.47	$\infty^{a)}$	$-\infty$	16.3	-75.5	23.1	-31.6

Under *Transformation B* the righthand side of (21) reduces to

$$r(c_r - c_m - c_d)\left(\frac{\gamma}{\lambda}\right)\left(\frac{\lambda+\gamma}{\lambda-\gamma}\right) Q_m/2 - rK_m\left(\frac{3}{2} - \frac{\gamma}{\lambda}\right) + rK_r/2.$$

The difference will be significant for large enough $|c_m + c_d - c_r|$, γ, and/or $|c_d|$ (see Table 3,4, and 5).

Table 4. Performance of $Q_m^{\overline{\text{AS}}}$ and Q_m^{AP} under the base-case scenario for $\Gamma = \gamma$ and various values of γ. [a] The objective function is an increasing function in Q_m

					Transform. A		Transform. B	
γ	Q_m^{AS}	$\tilde{AS}(Q_m^{\text{AS}})$	$Q_m^{\overline{\text{AS}}}$	$R(Q_m^{\overline{\text{AS}}})$	Q_m^{AP}	$R(Q_m^{\text{AP}})$	Q_m^{AP}	$R(Q_m^{\text{AP}})$
0	19.7	-20.67	20.0	0.0	20.0	0.0	20.0	0.0
5	24.5	-25.27	24.5	0.0	27.5	-0.6	32.1	-3.6
10	16.3	-25.00	16.3	0.0	23.1	-6.0	$\infty^{a)}$	$-\infty$
15	7.0	-28.66	7.1	0.0	12.1	-15.4	$\infty^{a)}$	$-\infty$
19	1.2	-33.42	1.2	0.0	2.1	-16.6	$\infty^{a)}$	$-\infty$

Table 5. Performance of $Q_m^{\overline{\text{AS}}}$ and Q_m^{AP} under the base-case scenario for $\Gamma = \gamma$ and various values of c_d. [a] The objective function is an increasing function in Q_m

					Transform. A		Transform. B	
c_d	Q_m^{AS}	$\tilde{AS}(Q_m^{\text{AS}})$	$Q_m^{\overline{\text{AS}}}$	$R(Q_m^{\overline{\text{AS}}})$	Q_m^{AP}	$R(Q_m^{\text{AP}})$	Q_m^{AP}	$R(Q_m^{\text{AP}})$
-15	16.3	-25.00	16.3	0.0	23.1	-6.0	11.5	- 6.1
-10	16.3	-25.00	16.3	0.0	23.1	-6.0	13.3	-2.1
-5	16.3	-25.00	16.3	0.0	23.1	-6.0	16.3	0.0
0	16.3	-25.00	16.3	0.0	23.1	-6.0	23.1	-6.0
5	16.3	-25.00	16.3	0.0	23.1	-6.0	$\infty^{a)}$	$-\infty$
10	16.3	-25.00	16.3	0.0	23.1	-6.0	$\infty^{a)}$	$-\infty$
15	16.3	-25.00	16.3	0.0	23.1	-6.0	$\infty^{a)}$	$-\infty$

7 Discussion

Although the net present value approach is the more appropriate framework, average cost models dominate the field of inventory control and production planning. In this paper we have shown that the traditional average cost approach, which does not make a distinction between opportunity costs of holding inventories and physical inventory costs, leads to reasonable results for single-source systems, but not necessarily for multi-source systems. The NPV approach does make a clear distinction between physical inventory costs and opportunity costs, since the two are not directly related. The latter does not depend on physical stocks at all, but only on the amount and timing of the investments.

The traditional approach only takes the opportunity costs of *holding inventories* into account, but this should not be a general rule. All cash-flows generate opportunity costs or yields that cannot be disregarded if the cash-flows depend on decision parameters. For example, in a manufacturing/remanufacturing system with disposal the throughput of the (re)manufacturing process is controlled by a decision variable. In that case also opportunity costs of set-ups and disposals should be taken into account. Clearly, these opportunity costs have got little to do with physical inventories.

Main conclusion of this paper is that basically there are two classes of models: a class for which a holding cost transformation exists that does not depend on decision variables, such that NPV coincides with AC (up to a constant), and a class for which such a transformation does not exist. A typical example of the latter class is a system with manufacturing, remanufacturing, and disposal.

Acknowledgement

The research presented in this paper makes up part of the research on re-use in the context of the EU sponsored TMR project REVersed LOGistics (ERB 4061 PL 97-5650) in which take part the Otto-von-Guericke Universitaet Magdeburg (D), the Erasmus University Rotterdam (NL), Eindhoven University of Technology (NL), the Aristoteles University of Thessaloniki (GR), the University of Piraeus (GR), and INSEAD (F).

Part of the research has been done during the first author's stay at the department of Technology Management of INSEAD, Fontainebleau, France, for which INSEAD is greatly acknowledged. The first author also acknowledges the financial support provided by the Dutch Organization for Scientific Research, NWO.

References

W. Beranek (1966). Financial implications of lot-size inventory models. *Management Science* 13(8):401-408.

M. Corbey, K. Inderfurth, E. van der Laan, S. Minner (1999). The Use of Accounting Information in Production and Inventory Control, Preprint 24/99 (FWW), University of Magdeburg, Germany.

R.W. Grubbström (1980). A principle for determining the correct capital costs of work-in-progress and inventory. *International Journal of Production Research* 18(2):259-271.

G. Hadley (1964). A comparison of order quantities computed using the average annual cost and the discounted cost. *Management Science* 10(3):472-476.

D.P. Heymann (1977). Optimal disposal policies for a single-item inventory system with returns. *Naval Research Logistics Quarterly* 24:385-405.

C. Hofmann (1998). Investments in modern production technology and the cash flow-oriented EPQ model. *International Journal of Production Economics* 54:193–206.

K. Inderfurth (1997). Simple optimal replenishment and disposal policies for a product recovery system with leadtimes. *OR Spektrum* 19:111-122.

Inderfurth K. and R. Teunter (1998). The 'right' holding cost rates in average cost inventory models with reverse logistics", Preprint nr. 28/98, Fakultät für Wirtschaftswissenschaft, Otto von Guericke Universität, Magdeburg, Germany. Magdeburg, Germany.

Y.H. Kim, K.H. Chung and W.R. Wood (1984) A net present value framework for inventory analysis. *International Journal of Physical Distribution & Materials Management* 14(6):68–76.

W.K. Klein Haneveld and R.H. Teunter (1998). Effects of discounting and demand rate variability on the EOQ, *International Journal of Production Economics*, 54:173–192.

E. Luciano and L. Peccati (1999). Some basic problems in inventory theory: the financial perspective, *European Journal of Operational Research*, 114:294-303.

K. Richter (1996). The EOQ repair and waste disposal model with variable setup numbers. *European Journal of Operational Research*, 95:313–324.

D.A. Schrady (1967). A deterministic inventory model for repairable items. *Naval Research Logistics Quarterly*, 14:391–398.

V.P. Simpson (1978). Optimum solution structure for a repairable inventory problem. *Operations Research* 26:270–281.

R. Teunter (1998). Economic ordering quantities for remanufacturable item inventory systems. Preprint nr. 31/98, Fakultät für Wirtschaftswissenschaft, Otto von Guericke Universität, Magdeburg, Germany.

H. E. Thompson (1975). Inventory Management and capital budgeting: a pedagogical note, *Decision Sciences* 6:383-398.

R. R. Trippi and D. E. Lewin (1974). A present value formulation of the classical EOQ problem, *Decision Sciences* 5:30-35.

E. van der Laan and M. Salomon (1997). Production planning and inventory control with remanufacturing and disposal. *European Journal of Operational Research* 102:264-278.

E. van der Laan, M. Salomon, R. Dekker and L. Van Wassenhove (1998). Inventory control in hybrid systems with remanufacturing, *Management Science* 45(5):733-747.

Appendix

If $\Gamma = 0$ there are no cash-flows related to remanufacturing operations. Hence, expressions (15) – (17) are given as

$$AS = p\lambda - r\left(\frac{K_m + Q_m c_m + \left(\frac{\gamma}{\lambda}\right) Q_m c_d \mathrm{e}^{-rQ_m/\lambda}}{1 - \mathrm{e}^{-rQ_m/\lambda}}\right),$$

$$\overline{AS} = p\lambda - c_m\lambda - c_d\gamma - \lambda K_m/Q_m - r(K_m + c_mQ_m)/2 + rc_d\left(\frac{\gamma}{\lambda}\right)Q_m/2,$$

and

$$AP = p\lambda - c_m\lambda - c_d\gamma - \lambda K_m/Q_m - h_mQ_m/2 - h_n\left(\frac{\gamma}{\lambda}\right)Q_m/2.$$

Safety Stocks in Capacity-constrained Production Systems

Michael Wagner

University of Augsburg, Department for Production and Logistics,
Universitätsstr. 16, 86135 Augsburg

Abstract. We present a simple calculation scheme for the determination of safety stocks for a capacitated single-stage multi-product production-inventory system. Demand is assumed to be stochastic and a given fill rate has to be fulfilled. The lotsizing and sequencing decisions are made on a rolling horizon basis considering sequence-dependent setup cost and time. Results from a simulation experiment show the reliability of the approach under varying parameter settings.

1 Introduction

More and more companies are using Advanced Planning Systems (APS) to plan and control the operations of their production/distribution network. These software tools support specific planning tasks with quite sophisticated modeling facilities and optimization algorithms. Most of the APS implementations are based on the concept of hierarchical planning, using a rolling horizon and deterministic models.[1] Therefore, in make-to-stock industries (we focus on the consumer goods industry) the whole planning process is based on more or less uncertain forecasts. Additional safety buffers (inventory or time) are necessary to guarantee a certain level of service for the customers.

In the following it will be shown how time-independent safety stocks SS_i can be calculated for products $(i, j = 1, \ldots, N)$, which are supplied by a single-stage capacitated production process. This multi-product process can be characterized by high sequence-dependent setup costs s_{ij} and times st_{ij} and thus requires simultaneous optimization of production lotsize and sequence. We assume that this lotsizing and sequencing problem (LSP) is based on net requirements m_{it}, which can be calculated from predetermined forecasts f_{it}. For the LSP a multitude of different MIP-formulations and solution procedures exists[2].

In this paper we will use the General Lotsizing and Scheduling Problem (GLSP[3]) to model the production planning process, but the results derived here are also applicable to similar LSP models. The time structure of the GLSP consists of macro periods t, which are used to model the inventory

[1] cf. Zijm (2000) and Stadtler and Kilger (2000)
[2] cf. Drexl and Kimms (1997)
[3] Meyr (1999) and Meyr (2000)

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

status, and micro periods s, which represent production lots and setup times. These micro periods are assigned to one specific macro period ($s \in S_t$)[4] and their length is variable. The planned inflow to the stock pile of product i in period t is then given by the sum of all micro period production quantities x_{is}: $y_{it} = \sum_{s \in S_t} x_{is}$. The objective function of the GLSP includes changeover costs and also the costs for cycle stock due to lotsizing. Solutions are usually generated by heuristic procedures, which are able to solve this hard problem in reasonable time.

Safety stocks SS_i are considered as minimum stock levels in the calculation of net requirements for the GLSP. If the production plan is frozen for some periods, then the planned safety stock has to cover the uncertainty in the frozen horizon and the time until the product is setup next (lead time). Therefore, one could calculate the necessary safety stock after solving the production planning problem, because the lead time is known then. However, the net requirements are based on safety stocks which have to be available in advance. Thus, in our model the calculation of safety stocks is considered to be a tactical problem, which is fixed in the operational production scheduling decision.

The safety stock level depends on the length of the frozen horizon and the replanning interval (= 1 / planning frequency). Frequent replanning based on updated forecasts increases the service level, because it is possible to respond to higher demand (e.g. by earlier production) and to incorporate forecast changes previously. But high replanning frequencies can also lead to more frequent setups and therefore to increased changeover effort. Furthermore, planning nervousness due to fluctuating forecasts is possible. Thus, it is necessary to find a suitable planning frequency balancing the advantages and disadvantages. The following three control mechanisms can be differentiated:

- permanent replanning: Changes of any kind initiate the replanning.
- fixed planning rhythm: A new plan is created each τ periods; in the meantime it is not possible to react to changes.
- event-based planning: The plan is revised every time a specific event occurs. Examples are the following:
  - The inventory position drops below a specific value (according to the reorder point in stochastic inventory control) or
  - backorders, which are expected in the following x periods.

The integration of a fixed planning rhythm and event-based replanning is possible. Then a fixed planning interval of τ periods is established, but events can require replanning inbetween two "regular" planning events.

The decision on the control mechanism (or the replanning interval) influences the necessary amount of safety stock. Increased responsiveness is possible due to frequent replanning, which reduces the risk interval, that has

[4] S_t denotes the set of micro periods assigned to macro period t

to be protected by safety stocks. Thus, if the control mechanism is not predetermined by company specific structures, both planning tasks have to be tackled simultaneously (under consideration of dependencies).

In the following section a short overview on literature concerning safety stocks in production systems with rolling horizons is given (Section 2). Subsequently, in Section 3 a model for the calculation of safety stocks is introduced. Some numerical examples show the efficiency of the proposed concept in Section 4. In Section 5 we give some concluding remarks and an outlook on further research.

2 Related Literature

The calculation of safety stocks for rolling horizon planning systems has been discussed intensively for MRP-systems (material requirements planning).[5] The MRP planning concept differs in a few issues from the APS concept considered here. In contrast to MRP-systems the APS-tool considers finite capacities and multi-product lotsizing. Furthermore, it integrates the decision on the sequence because of the sequence dependency of changeovers. Most authors restrict their research to safety stocks for finished products and thus to the master production scheduling (MPS) level. A good review of available concepts (safety stocks, safety time, hedging) for the treatment of uncertainty in multi-level MRP-systems is given by Wijngaard and Wortmann (1985). Nevertheless, they assume fixed lead times and lot-for-lot ordering in all models.

The cost impact of stochastic demand in MRP lot-sizing models is analyzed by de Bodt and van Wassenhove (1983). The calculation is based on the time between two successive setups (time between orders (TBO) = natural cycle). Changes in demand or forecast cannot cause replanning in this case, because the next setup is fixed by the lotsize. They validate their model by simulation and report some dynamics for this single-stage uncapacitated MRP system.

A similar study was conducted by Wemmerlöv and Whybark (1984): They compare heuristic lotsizing methods for deterministic and stochastic demand and conclude, that the relative advantage of specific heuristics decreases in the stochastic case.

Planning of safety stocks for finished products follows exclusively the classical MRP literature and therefore is based on the assumption of fixed lead times for each product. In contrast to this philosophy Blackburn et al. (1987) study the influence of various concepts on the reduction of uncertainty in multi-stage MRP-systems. They compare safety stocks, fixation of orders in the replanning interval and lot-for-lot ordering for components. For the specific case of assembly structures Lambrecht (1984) develops a heuristic for

[5] Wijngaard and Wortmann (1985)

safety stock calculation based on the Clark-Scarf approach. They use deterministic lead times, periodic review (s,S)-ordering and centralized control.

While the above mentioned literature does not focus on the rolling horizon concept, the following summarizes the research on the problem of setting parameters (e.g. replanning interval, planning horizon etc.) for this concept and their impact on safety stocks. Yano and Carlson (1987) explain the trade-offs between rescheduling and safety stocks. They consider 2-echelon production systems, which are controlled by using a rolling MRP-plan, and they differentiate between two scenarios: One in which the production orders are fixed during a frozen horizon and one in which all orders can be rescheduled each period. The calculation of safety stocks for the first scenario is based on a periodic review policy using the *natural cycle* as review period. Their simulation results show, that it is not possible to determine which concept is better a priori. In the end they state, that frequent rescheduling should be done with caution as costs increase due to this. Lin and Krajewski (1992) also investigate the single-stage case (MPS) without capacity restrictions. They develop a detailed concept for the determination of replanning interval, frozen interval and forecast window. The calculation of safety stocks is based on the standard deviation of forecast errors in the frozen interval. The proposed concept is based on analytical relationships among the parameters, which are derived by the authors. They verify their findings according to cost and service level by simulation of the planning system. Sridharan et al. (1988) focus on the MPS schedule stability in Wagner-Whitin schedules. They vary the parameters planning horizon, replanning interval and frozen period. The frozen interval is either given by a specific time interval or by a certain number of future orders.

The first paper (to our knowledge), which incorporates capacities in these models has been published by Yang and Jacobs (1999). They show that replanning interval and planning horizon have very little impact on the performance of a capacitated job-shop. But the influence of the dispatching rule is quite large if the forecast error is lumpy. Metters and Vargas (1999) extend the rolling horizon concept by adding reorder points for single products. They recalculate the net demands in each period, but the quantities are only changed, if the inventory position falls below the reorder point. Therefore, the planning concept is event-based and the schedule can react to demand changes in each period. Since this model does not consider restricted capacities and multiple products, it cannot be applied to the APS-case investigated here, because rescheduling of one product in the capacitated case causes changes in the lead time of other products too.

The concepts from MRP-literature presented here provide some general ideas, which can also be applied to APS environments. However, almost none of the safety stock calculations consider capacity restrictions. Having significant influence on the lead time the restricted capacity has to be incorporated in those models. The following section shows how to deal with this additional requirement.

3 Model Description

We study a single-stage, multi-product, capacitated production-inventory system with stochastic demand and lotsizing. The system has to ensure a given fill-rate constraint while minimizing the costs for setups, cycle stock (stock due to lotsizing) and safety stock. All unsatisfied demand will be back-ordered. The problem is decomposed in a tactical safety stock problem and the operational lotsizing and scheduling problem. This enables us to model the LSP as deterministic optimization problem and solve it by one of the proven solution procedures (see e.g. Meyr (1999)).

3.1 System Dynamics

For the safety stock model (SSM) the detailed timing of planning decisions and the corresponding system dynamics have to be known: We use a rolling planning concept with a fixed planning rhythm as it is shown in Figure 1. This means, that a production schedule is created every τth period (replanning interval) for the next T periods (planning horizon). Only the first τ periods are frozen and have to be put into practice. In our model the frozen horizon equals the replanning interval, whereas in practical applications the frozen horizon may be longer than the replanning interval.

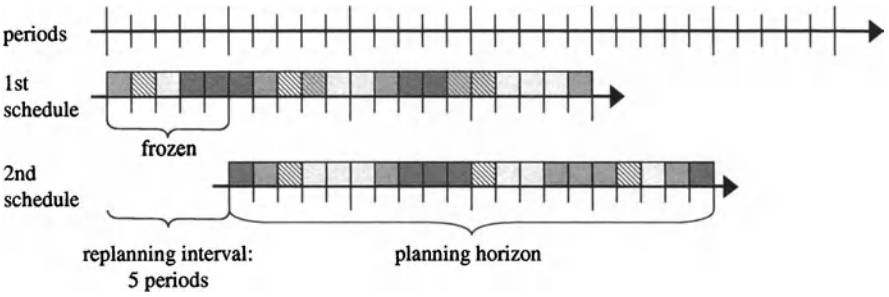


Fig. 1. Rolling planning concept

The following notation will be used below:

p	beginning of the planning horizon ($p \in \{0, \tau, 2\tau, \ldots\}$)
d_{it}	demand for product i in period t
D_i	random variable of the period demand of product i
$\overline{d}_i, \sigma_i$	mean and standard deviation of D_i
f_{it}	demand forecasts for product i in period t
m_{it}	net requirements for product i in period t
SS_i	planned safety stocks for product i

y_{it}	scheduled production of product i in period t
TBO_i	average production cycle of product i (time between two consecutive setups)
l_{ip}	planned lead time of product i according to schedule created in p
L_i	random variable of the lead time of product i
I_{it}	projected net stock of product i at the end of period t for $t = p+1, \ldots, p+T$
A_i	physical stock level at the beginning of the planning cycle
B_i	backordered demand at the beginning of the planning cycle
I_{ip}	initial net stock $= A_i - B_i$

At the beginning of a planning cycle ($t = p$) net requirements for the planning horizon T are calculated from initial inventories, backorders, planned safety stocks and demand forecasts. These are used as input to the LSP which returns a production schedule for the next T periods. We assume, that lots for a single product i span one ore more consecutive periods t for which the production quantities $y_{it} > 0$.
The net requirements are calculated according to equations (1):

$$\begin{aligned} m_{it} &= \max\{0, f_{it} + SS_i - I_{i,t-1}\} \\ I_{it} &= \max\{SS_i, I_{i,t-1} - f_{it}\} \end{aligned} \quad ,\text{if} \quad t = p+1, \ldots p+T \qquad (1)$$

The definition of the net requirements implies that initial backorders have to be produced in the first period of the planning horizon $p + 1$. This also applies to the case of low safety stocks, which have to be refilled in the same manner. Table 1 shows the results of the above procedure for a simple example with four products. The planning procedure starts at the end of period $p = 0$ and comprises a planning horizon of $T = 10$ periods. The net requirements include predefined safety stocks of $SS_i = 10$ units for all products i.

From the definition of m_{it} it is obvious, that there exist three different types of requirements: forecast, safety stock and backordered demand. All types are summed up in one figure and thus have the same priority for production. This restricts the flexibility of the system and should be tackled in further research.

3.2 The Safety Stock Model

Our concept is based on well-known periodic review inventory policies. The *review period* (replenishment cycle) is equivalent to the replanning interval τ and the supply is given by planned production lots. However, in contrast to classical inventory policies no (production) orders are generated by using simple 'order-up-to'-rules. Instead, the timing and size of production lots (orders) is optimized in a separate lotsizing and scheduling step.

Table 1. Calculation of net requirements and production schedule for a simple example

	period										
product 1	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
forecast f_{1t}		10	12	13	20	25	15	19	17	14	13
project. net stock I_{1t}	-10	10	10	10	10	10	10	10	10	10	10
net requirement m_{1t}		30	12	13	20	25	15	19	17	14	13
scheduled production y_{1t}		44.1	50.0	5.9			18.7	32.8		13.5	13.0
		lead time = 1									
product 2	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
forecast f_{2t}		21	18	22	19	17	13	17	19	20	16
project. net stock I_{2t}	50	29	11	10	10	10	10	10	10	10	10
net requirement m_{2t}		0	0	21	19	17	13	17	19	20	16
scheduled production y_{2t}				43.5	43.5				19.0	36.0	
		lead time = 3									
product 3	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
forecast f_{3t}		16	18	14	18	17	10	17	19	14	15
project. net stock I_{3t}	70	54	36	22	10	10	10	10	10	10	10
net requirement m_{3t}		0	0	0	6	17	10	17	19	14	15
scheduled production y_{3t}					6.0	46.2	30.8				15.0
		lead time = 4									
product 4	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
forecast f_{4t}		12	15	17	15	18	11	16	15	18	19
project. net stock I_{4t}	100	88	73	56	41	23	12	10	10	10	10
net requirement m_{4t}		0	0	0	0	0	0	14	15	18	19
scheduled production y_{4t}								16.6	30.4		19.0
		lead time = 7									

The production *lead time* will be defined as the interval between the planning period p and the planned start time of the first production lot:

$$l_{ip} := \min_{p<t\leq T}\{t|y_{it} > 0\} - p$$

In the example in Table 1 the resulting lead times are $l_{1p} = 1, l_{2p} = 3, l_{3p} = 4, l_{4p} = 7$ periods. Besides the length of the replanning interval τ, the timing of the production quantities y_{it} mainly influences the lead times. As these production quantities are underlying a complex optimization procedure it is not possible to give an analytical expression for y_{it}. Therefore, we are modelling the lead times l_{ip} as discrete stochastic variables L_i, which are drawn from an empirical distribution.

PROPOSITION: The distribution of the lead time L_i is *independent* of the safety stock level SS_i.

PROOF: The safety stock SS_i only influences the lead time l_{ip}, if the production quantities y_{it} depend on the safety stock. The production schedule is created from net requirements m_{it}, which comprise 'refill quantities' for safety stocks, if the initial stock is below SS_i. If an initial backlog exists, then the requirements also contain 'refill quantities' for this backlog. The total refill quantity stays the same if safety stocks are changed, because the backlog is substituted by safety stocks and vice versa. Both components are equally weighted input to the net requirements calculation (1) and thus m_{it} are independent of safety stock levels.

This property allows us to draw samples from historical production schedules, calculate frequency distributions of the lead time and therefrom safety stocks for future operation of the system. The lead time distribution usually varies heavily depending on the main factors influencing the outcome of the lotsizing and scheduling decision. The most important are the finite production capacity and the sequence-dependent changeovers. For instance, reducing the capacity of the example in Table 1 by 10% results in the following lead times: $l_{1p} = 1, l_{2p} = 3, l_{3p} = 2, l_{4p} = 7$. In this case only the lead time of product 3 is changed from 4 to 2, but the variability of lead times increases significantly due to capacity reduction. If not enough historical data is available for the frequency distribution, then it might be necessary to draw the sample of lead times from a simulation of the production system.

In the following we consider the case of a fill-rate constraint (β) for the safety stock SS calculations.[6] The fill-rate is defined as follows[7]:

$$\beta = 1 - \frac{E(\text{shortage per replenishment cycle})}{E(\text{demand per cycle})} \tag{2}$$

The expected demand per cycle can easily be approximated by the mean demand $\bar{d}$ in TBO periods. However for the expected shortage per replenishment cycle (ESPRC) it is necessary to know the density functions $f_{D^R}(z^R)$ and $f_{D^L}(z^L)$ of the demand during the risk interval $R = L + \tau$ and the lead time L:

$$\begin{aligned} ESPRC = & \int_{SS+\bar{d}\cdot R}^{\infty} (z^R - SS - \bar{d} \cdot R) \cdot f_{D^R}(z^R) dz^R - \\ & - \int_{SS+\bar{d}\cdot R}^{\infty} (z^L - SS - \bar{d} \cdot R) \cdot f_{D^L}(z^L) dz^L \end{aligned} \tag{3}$$

[6] The index i is omitted in the following because the calculation is similar for all products.

[7] cf. Tempelmeier (1999), p. 370

The demand in the risk interval and the lead time depends on the stochastic variables D (demand per period) and R (length of the risk interval) or L (lead time). Therefore, the distribution function of the stochastic variables D^R and D^L can be derived by convolution of the demand variable and the corresponding time variable.

In the following we assume that the demand per period d_t is normally distributed[8] (i.i.d.) with mean $\bar{d}$ and standard deviation σ_d. The density function of the discrete random variable L, however, is calculated empirically from the frequency distribution of a sample. This is reasonable, because it is usually hard to find a standard distribution which approximates the frequency distribution quite well. In this case all potential values of the lead time are known $l_{min} \leq l \leq l_{max}$ and therefore the common density of D and R can be calculated by conditional probabilities (Eppen and Martin (1988)):

$$f_{D^R}(z^R) = \sum_{r=l_{min}+\tau}^{l_{max}+\tau} f_{D^R}(z^R|R=r) \cdot P(L=r-\tau)$$

$$f_{D^L}(z^L) = \sum_{l=l_{min}}^{l_{max}} f_{D^L}(z^L|L=l) \cdot P(L=l)$$

The expected shortage per cycle can be expressed by the following equation (Tyworth (1992)):

$$ESPRC = \sum_{l=l_{min}}^{l_{max}} ESPRC_l \cdot P(L=l) \tag{4}$$

Where the expected shortage for a specific duration of the lead time[9]

$$ESPRC_l = G_u\left(\frac{SS}{\sigma_{l+\tau}}\right) \cdot \sigma_{l+\tau} - G_u\left(\frac{SS+\bar{d}\cdot\tau}{\sigma_l}\right) \cdot \sigma_l$$

can be calculated using Brown's rational [10] $G_u(k)$. The standard deviation in x periods σ_x is based on the standard deviation of forecast errors per period σ_e ($\sigma_x = \sqrt{x} \cdot \sigma_e$).

If a specific service level has to be guaranteed, then the expected shortage per cycle should not exceed the maximum shortage $(1-\beta) \cdot \bar{d} \cdot TBO$:

$$(1-\beta) \cdot \bar{d} \cdot TBO \geq \sum_{l=l_{min}}^{l_{max}} ESPRC_l \cdot P(L=l) \tag{5}$$

As the inequality (5) cannot be transformed for SS, the minimum safety stock has to be evaluated by a numerical search procedure (e.g. Newton-Raphson, bisection etc.) for each product. Although the safety stock problem

[8] The procedure is also applicable to other standard distributions.

[9] see de Kok (1990)

[10] for the derivation of $G_u(k)$ see Silver et al. (1998), Appendix B for example

was decomposed in single product inventory models, the influence of the common capacity can be captured by stochastic lead times. The major drawback of the model is the assumption, that historical data is available to derive the frequency distribution of lead times. Analytical approximations would be favourable, but the production problem is too complex to gain enough insight into the dynamics of the system. Furthermore the production problem has to be solved by a heuristic procedure, which possibly returns different schedules depending on the parameter setting and the allowable computation time. The safety stock procedure can be summarized by the following steps:

1. Determine the replanning interval τ and the mean demand $\bar{d}$.
2. Calculate the standard deviation of forecast errors σ_e from historical demand and forecast data and derive a frequency distribution of lead times (for each product i) from historical production schedules.
3. Determine the safety stock levels SS for each product i by numerical search.

4 Case Study and Simulation Results

In the following the steps shown above will be explained by using a simple example. We consider a production line which produces 4 different sku's to stock. All products require the same capacity amount per item produced and customer demand is normally distributed with the parameters shown in Table 2.

Table 2. Parameters for the demand distribution

products $i =$	1	2	3	4
$\bar{d}$	100	80	120	60
σ_d	10	20	30	20

The forecasts were generated by using exponential smoothing (see for example Silver et al. (1998)). The company reschedules according to a fixed planning rhythm of three weeks (= 15 days / periods) for a planning horizon of four weeks (= 20 periods). The changeover costs and inventory costs are preset such that each product is produced approximately once or twice a week.

The desired fill-rate is set to 98% for all products. The lotsizing and sequencing decision is met by a local search heuristic for the GLSP[11]. In the first scenario it is assumed that the capacity is not constrained and therefore backorders cannot cause infeasible production planning problems. However,

[11] General Lotsizing and Scheduling Problem, Meyr (1999) and Meyr (2000)

this is possible, if the capacity is constrained and the backorder level at the beginning of the planning horizon B_i creates net requirements $\sum_i m_{it}$ (+ changeover time) which exceed the available capacity although the mean demand does not. As the deterministic lotsizing and scheduling model has to fulfill all demands (hard constraint), it does not find a feasible solution to the planning problem. This means, that requirements have to be delayed to later periods to get a feasible solution. The following (heuristic) procedure can be used to meet the required service level anyhow and to minimize additional cost where possible:

1. Postpone net requirements which are necessary to refill the safety stocks SS_i until a feasible plan is reached. If this is not possible, then "planned" shortages have to be taken in a second step:
2. Postpone forecasted demand f_{it} in following periods for products which have no initial backorders B_i. This step eventually has to be repeated for all products until a feasible plan can be determined.

First we consider an example with infinite capacity. In this case the lead time distribution in Table 3 and Figure 2 could be calculated from historical data.[12]

Table 3. Distribution of lead times in the uncapacitated case

$l =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$P(L_1 = l)$	0.74	0.16	0.09	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$P(L_2 = l)$	0.34	0.07	0.09	0.09	0.16	0.10	0.10	0.04	0.00	0.01	0.00	0.00	0.00	0.00	0.00
$P(L_3 = l)$	0.65	0.14	0.12	0.06	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$P(L_4 = l)$	0.29	0.06	0.05	0.12	0.16	0.06	0.10	0.07	0.04	0.01	0.02	0.01	0.01	0.00	0.01

For each of the items the minimum safety stock has to be found by the Newton-Raphson method. Considering the mean demand (refer to Table 2) the maximum shortage per replenishment cycle (see Table 4) can be calculated. The minimum safety stock SS is then derived by equation (5).

Table 4. Safety stocks for $\beta = 98.0\%$

product	**1**	**2**	**3**	**4**
max. shortage per replenishment cycle	558	519	595	389
safety stock SS	11.7	70.9	115.8	92.5
simulated fill-rate [%]	98.0	98.6	98.0	98.4

[12] In our case the lead times were generated by simulation.

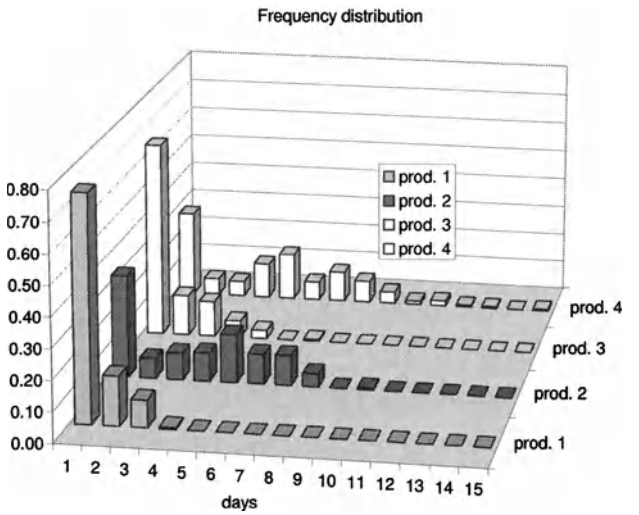


Fig. 2. Distribution of lead times (uncapacitated case)

The validity of the model was tested by an extensive simulation study with different replanning intervals, demand parameters, cost parameters and service levels. For the example shown above the fill-rates given in Table 4 were measured in the simulation. The duration of the simulation was 3000 periods. In all simulated cases the deviation from the given fill-rate was beyond 1%. But if mean lead times are used instead of the real distributions, then the deviations rise significantly compared to the proposed method.

This difference between the two models is increasing, if constrained capacities have to be considered. For example, if the mean utilization is about 80% in the case described above, then the setup cycles change significantly and thus also the lead time distribution (see Table 5 and Figure 3).

Table 5. Distribution of lead times in the capacitated case

$l =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$P(L_1 = l)$	0.53	0.33	0.11	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$P(L_2 = l)$	0.61	0.10	0.12	0.07	0.04	0.04	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$P(L_3 = l)$	0.56	0.24	0.12	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$P(L_3 = l)$	0.53	0.09	0.07	0.09	0.05	0.03	0.06	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01

In this case shorter lead times leading to lower safety stocks can be observed. But on the other hand more setups are necessary and the costs for

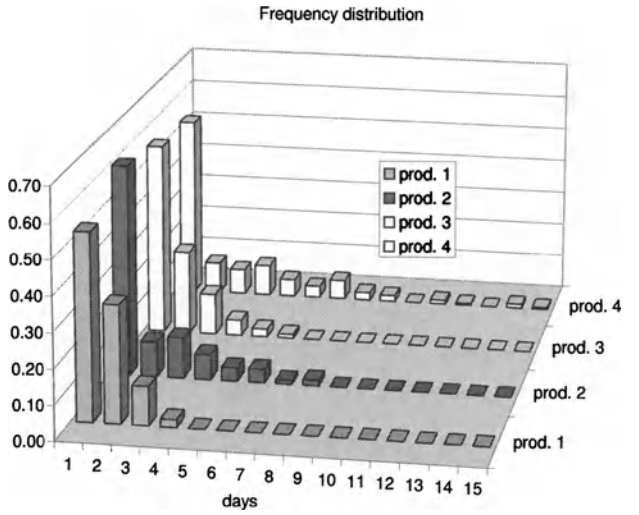


Fig. 3. Distribution of lead times (capacitated case)

changeovers increase. In our case the increase in changeovers outweighs the reduction in inventory cost and the total costs rise by 13 % (see Table 6).

Table 6. Cost comparison

	uncapacitated	**capacitated**	**change**
stock cost	3,331,578	3,003,469	-10%
changeover cost	2,543,300	3,642,600	+43%
total	5,874,878	6,646,069	+13%

In Section 1 we already pointed out the dependency of the replanning interval and the level of safety stocks. If both parameters are variable and can be optimized, then the question of the right objective function arises. In the previous Section the minimum safety stock was calculated under consideration of a specific service level and the replanning interval. But if the replanning interval is variable, then the stock which has to be optimized also comprises the lotsizing or cycle stock. Both stock types can be reduced by shorter replanning cycles, but at the same time the number of setups is increased. Thus, the objective function has to include stock costs for both stock types and also changeover costs. In this objective function all relevant performance measures are captured and therefore it is not necessary to measure

the planning stability or the planning nervousness[13], which is already covered by changeover costs.

5 Conclusion and Outlook

Although, the Advanced Planning Systems are using deterministic algorithms for lotsizing and scheduling, they lack support for the calculation of safety buffers in production systems in most cases. This either results in frequent plan revisions (and therefore nervousness) or the adaption of safety stocks or times which are based on "experience" or "rules of thumb". Our concept applies standard methods to this complex task and provides fast and robust access to reliable safety stock settings. As the APS are implemented in an rolling horizon concept in most cases, the method can be used for many capacity-constrained multi-product production systems.

More flexibility can be expected from event-based planning systems. These systems can react faster to changing demand or forecast situations. Our concept is not directly applicable to these environments, because the event which causes replanning is triggered by only one product but affects also all other products on a machine. This fact has to be integrated into the model explicitly.

In the consumer goods industries products are often produced at one site but shipped to more than one distribution center. The extension of our model to those 2-echelon divergent structures is possible if the stock points are controlled locally. Under certain circumstances the distribution of lead times for the second echelon can even be derived analytically[14].

References

Blackburn, J.D., Kropp, D.H., Millen, R.A. (1987) *Alternative approaches to schedule instability: a comparative analysis*, International Journal of Production Research, Vol. 25, No. 12, 1739–1749

de Bodt, M.A., van Wassenhove, L.N. (1983) *Cost increases due to demand uncertainty in MRP lot-sizing*, Decision Sciences, Vol. 14, 345-362

Drexl, A., Kimms, A. (1997) *Lot sizing and scheduling – survey and extensions*, European Journal of Operational Research, Vol. 99, No. 2, 221–235

Eppen, G., Martin, R. (1988) *Determining safety stocks in the presence of stochastic lead time and demand*, Management Science, Vol. 34, No. 11, 1380-1390

Kok, A.G. de (1990) *Hierarchical production planning for consumer goods*, European Journal of Operational Research, Vol. 45, 55–69

Lambrecht, M.R., Muckstadt, J.A., Luyten, R. (1984) *Protective stocks in multi-stage production systems*, International Journal of Production Research, Vol. 22, No. 6, 1001–1025

[13] see Blackburn et al. (1987), Sridharan et al. (1988) and Yano and Carlson (1987)

[14] Tempelmeier (2000)

Lin, N., Krajewski, L. (1992) *A model for master production scheduling in uncertain environments*, Decision Sciences, Vol. 23, No. 4, 839–860

Metters, R., Vargas, V. (1999) *A comparison of production scheduling policies on costs, service level, and schedule changes*, Production and Operations Management, Vol. 8, No. 1, 76–91

Meyr, H. (1999) *Simultane Losgrößen- und Reihenfolgeplanung für kontinuierliche Produktionslinien*, Gabler, Wiesbaden

Meyr, H. (2000) *Simultaneous lotsizing and scheduling by combining local search with dual reoptimization*, European Journal of Operational Research, Vol. 120, No. 2, 311–326

Silver, E.A., Pyke, D.F., Peterson, R. (1998) *Inventory Management and Production Planning and Scheduling*, 3. Aufl., John Wiley & Sons, New York

Sridharan, S.V., Berry, W.L., Udayabhanu, V. (1988) *Measuring master production schedule stability under rolling planning horizons*, Decision Sciences, Vol. 19, No. 1, 147–166

Stadtler, H., Kilger, C. (Hrsg.) (2000) *Supply Chain Management and Advanced Planning*, Springer, Berlin

Tempelmeier, H. (1999) *Material-Logistik*, 4. Aufl., Springer, Berlin

Tempelmeier, H. (2000) *Inventory service-levels in the customer supply chain*, OR Spektrum, Vol. 22, No. 3, 361–380

Tyworth, J.E. (1992) *Modeling transportation-inventory trade-offs in a stochastic setting*, Journal of Business Logistics, Vol. 13, No. 2, 97–124

Wemmerlöv, U., Whybark, D. (1984) *Lot-sizing under uncertainty in rolling schedule environment*, International Journal of Production Research, Vol. 22, No. 3, 467–484

Wijngaard, J., Wortmann, J. (1985) *MRP and inventories*, European Journal of Operational Research, Vol. 20, 281–293

Yano, C., Carlson, R. (1987) *Interaction between frequency of rescheduling and the role of safety stock in material requirements planning systems*, International Journal of Production Research, Vol. 25, 221–232

Yang, K.K., Jacobs, F.R. (1999) *Replanning the master production schedule for a capacity-constrained job-shop*, Decision Sciences, Vol. 30, No. 3, 719–748

Zijm, W.H.M. (2000) *Towards intelligent manufacturing planning and control systems*, OR Spektrum, Vol. 22, No. 3, 313–345

Approximations for the Waiting Time in (s,nQ)-Inventory Models for Different Types of Consolidation Policies

S.R. Smits and A.G. de Kok

Faculty of Technology Management
Technische Universiteit Eindhoven
Pav E6, P.O. Box 513,
5600 MB Eindhoven
The Netherlands

Abstract. In many practical situations the coordination of transportation management and inventory management may lead to considerable cost reductions. Transportation management includes the application of different types of shipment consolidation policies. Shipment consolidation takes into account the logistics strategy of combining two or more shipment orders to optimize transportation. When the shipment consolidation policy changes, the shipment lead time changes as well and if the lead time changes, the inventory policy needs to be re-evaluated, since changing lead times affect customer service. In this paper the lead time comprises two elements: waiting time due to order consolidation and the shipment time. The lead time is an important parameter for inventory management. We derive approximations for the lead time behaviour in (s,nQ) models where the items are consolidated according to different types of consolidation policies.

1 Introduction

The coordination of inventory management and transportation management is crucial for an efficient management of the supply chain. Transportation management includes the application of different types of shipment consolidation policies. The consolidation policy coordinates shipment processes of different item orders for the same (intermediate) destination, and this can lead to a reduction in transportation costs. Higginson and Bookbinder (1994) distinguish between two types of consolidation policies: the time policy and the quantity policy. The time policy dispatches orders when a shipping date is expired. The shipping date is usually set through consideration of service levels. Higginson and Bookbinder (1995a) give some normative approaches to set the shipping date. The quantity policy dispatches orders when a fixed quantity is consolidated. Higginson and Bookbinder (1995b) use a Markov chain model to determine the optimal consolidation policy given an (s,S) inventory policy generating shipment orders.

Another line of research is the joint replenishment or coordinated replenishment problem. Goyal and Satir (1989) present an early review of all

A. Klose et al. (eds.), *Quantitative Approaches to Distribution Logistics and Supply Chain Management*

models, starting from a simple deterministic problem. In the joint replenishment literature we observe two types of control policy; viz. the continuous review can-order policy (s_i,c_i,S_i) and the periodic review order-up-to policy (R_i,S_i). In the continuous can-order policy (s_i,c_i,S_i), when the inventory position of an item i reaches the must-order point s_i, a replenishment is triggered as to raise the item's inventory position to order-up-to level S_i. Meanwhile, any other item in the group with an inventory position at or below its can-order point c_i is included in the replenishment as to raise the inventory position up to S_i. See, e.g., Liu and Yuan (2000) or Federgruen, Groenevelt and Tijms (1984). In the periodic review (R_i,S_i) policy, the inventory position of item i is inspected with intervals R_i and the review moments are coordinated in order to consolidate orders of individual items, see Viswanathan (1998).

In this paper, we analyze shipment consolidation policies under the assumption of compound renewal customer demand. The compound renewal demand process enables accurate modeling of real-world demand processes. In the literature discussed above order-up-to-policies are employed for inventory management. In practice, it is often more appropriate to employ (s,nQ)-policies, which take into account restrictions imposed by material handling units such as pallets and boxes. The focus in this paper is to model the interaction between shipment consolidation processes and inventory management policies. In general this involves multiple items or stock keeping units (sku's) and multiple stock locations. It is easy to see that a building block for the analysis of the interaction between shipment consolidation and inventory management, is the analysis of a line haul between two stocking locations, for example between a warehouse and a retailer. We assume for ease of reference that the warehouse holds stocks of multiple items. The same items are held by the retailer who sells to customers. The inventories at the retailer are controlled according to an (s,nQ)-policy. The retailer has to satisfy fill rate requirements for all items. The reorder level, which ensures the required fill rate, depends on the lead time of orders from the warehouse to the retailer. The lead time of an order comprises the waiting time for truck departure and the transportation time. The waiting time for truck departure is the main subject of this paper. The contribution of this paper to literature is twofold. First of all, we analyze shipment consolidation policies under the assumption of compound renewal demand, where this demand represents the customer orders at the retailer. We present an overall analysis of this problem integrating shipment consolidation and inventory management taking into account material handling restrictions. The latter is dealt with through the use of (s,nQ)-policies.

In this paper, we use the method of Whitt (1982) to superpose renewal processes with mixed-Erlang distributed inter-renewal times. Notice that the superposed process is not a compound renewal process. Yet, our analysis reveals that in line with the research of Whitt (1982), assuming that the

superposed process is renewal yields good approximations for performance characteristics, c.f. Smits, de Kok and, van Laarhoven (2000).

The sequel of this paper is organized as follows: In section 2, we describe in detail the model and we derive approximations for the waiting time. In section 3, we test the approximations through extensive computer simulations and in the last section, we give some conclusions and indicate a few thoughts for further research.

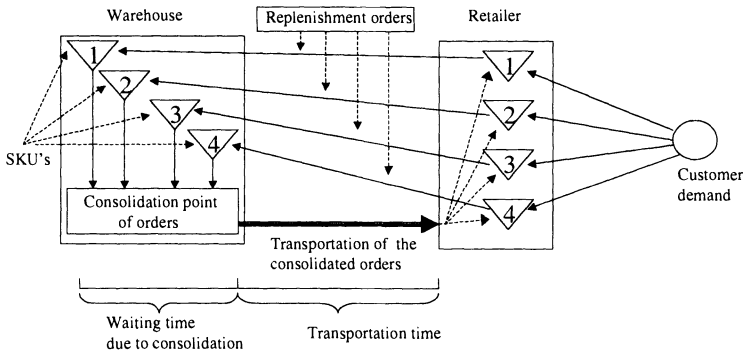


Fig. 1. Schematic representation of the replenishment process

2 Model Description

The model considers a line haul between a warehouse and a retailer. At both locations stocks of all items are kept. At the retailer the customer demand for each item arrives according to a compound renewal process, i.e., customer orders for an item arrive according to a renewal process and the demand per customer has some arbitrary distribution function. Demands of different customers for an item are independent and identically distributed. We furthermore assume that the compound renewal demand processes for different items are independent of each other. Additional constraints of the model are the item fill rate constraints at the retailer. This implies that at

the retailer for each item a target fill rate is given. The fill rate is the fraction of demand directly delivered from shelf. The inventories at the retailer are controlled by (s,nQ)-policies. It operates as follows: as soon as the inventory position, which is expressed as the physical inventory plus the stock on order minus the backorders, drops below reorder level s an amount nQ is ordered such that the inventory position is raised to a value between s and $s+Q$. Q is called the batchsize, n is an integer. The demand that cannot be met immediately is backordered. We assume that the warehouse always has enough stock to fulfill the replenishment orders towards the retailer. To be able to calculate the reorder levels of the different items, the lead time towards the retailer is needed. In our model the lead time comprises the waiting time due to transport consolidation and the handling and transportation time. In this paper, we derive approximations for the waiting time due to consolidation for the time policy and the quantity policy. Figure 1 gives a graphical representation of the model. Below a list of the used notation is introduced.

Parameters and variables

L_d	driving time
Z	waiting time due to shipment consolidation
L^*	lead time
T	time between two truck departures. In the time policy T is deterministic and in the quantity policy T is stochastic
A_i	time between two subsequent arrivals of item i at the retailer
D_i	demand size of item order i (in volume) at the retailer
P_{2i}	fill rate at the retailer
$D_i(L^*)$	demand for item i during the lead time
s_i	reorder level of item i (in volume)
X_i^+	physical inventory level at an arbitrary point in time
Q_{max}	predetermined consolidation quantity (in volume)
$\Delta(t)$	consolidated quantity at time t, $\Delta(T)$ is the shipped quantity
Q_i	batchsize of item i (in volume)
O_i	order process of the retailer towards the warehouse for item i
R_i	time between order placements at the warehouse for item i
O^*	aggregate order process at the warehouse
$O^*_{\neq i}$	aggregate order process of all item except i at the warehouse
R^*	aggregate process of the time between two order placements at the warehouse
$R^*_{\neq i}$	aggregate process of the time between two order placements for all items except i at the warehouse
V	rest part of the split order in the quantity policy with partial shipments

U undershoot process
$N(T)$ is defined as the number of arrivals in $(0,T]$
W_i is defined as $O^*_{\neq i} + V + O_i$
$N(X)$ is defined as the number of arrivals between the arrival of an arbitrary customer and the departure of the truck
$Y(t)$ inventory position at moment t

Functions and Operators
$E[Y]$ expectation of the random variable Y
$\sigma^2(Y)$ variance of the random variable Y
$E[Y^2]$ second moment of the random variable Y
c_Y coefficient of variation of the random variable Y
$(y)^+$ max(0,y)
$P\{A\}$ probability of event A
$F^{n*}(t)$ n-fold convolution of $F_y(t)$
$F_y(t)$ pdf of random variable Y
$M(t)$ renewal function, $M(t) = \sum_{n=0}^{\infty} F^{n*}(t)$ associated with pdf $F_y(t)$

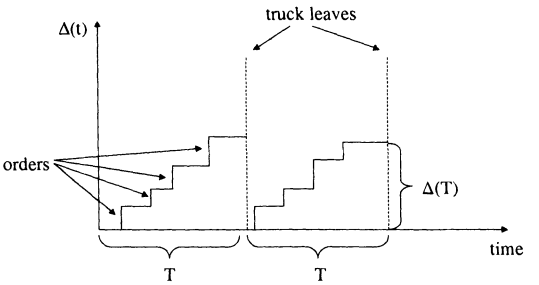


Fig. 2. Sample path of the collected quantity for the time policy

Time policy

In case consolidation employs a time policy the trucks depart at fixed time intervals T (for example, every week). Figure 2, gives a schematic representation of the time policy. All replenishment orders arriving within one time

interval T are consolidated and shipped together to the retailer. We define $\Delta(t)$ as the collected quantity at moment t and $\Delta(T)$ as the consolidated amount that is shipped. We want to find expressions for the time between the arrival of an arbitrary order and the departure of the truck.

Quantity policy

In case consolidation employs a quantity policy, we distinguish between two alternatives; partial shipments and full shipments/flexible truck capacity. In both cases we assume that at time 0 a truck leaves the warehouse.

i) ***Partial shipments***

The orders are consolidated until the collected quantity $\Delta(t)$, is larger than or equal to a predetermined quantity Q_{max}. The last order $O_{N(T)}$ is split such that $\Delta(T)$ is equal to Q_{max}, where T is the time of the first truck departure after time 0. The consolidated quantity $\Delta(T)$ is shipped to the retailer and the consolidation process starts all over again with as starting quantity the remaining part of the order V. So $O_{N(T)} - V$ leaves directly and the remaining part V leaves with the next truck. Figure 3a) gives a schematic representation of the process. In order to determine the probability distribution function of $\Delta(t)$ we apply the following proposition.

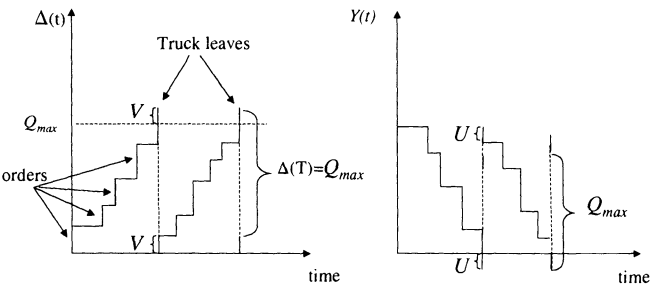


Fig. 3. Schematic representation of the similarities between the shipment consolidation policy with partial shipments and the $(0,nQ_{max})$ inventory policy

Proposition 1. *The consolidation process of a quantity policy with partial shipments under a compound renewal demand process is equivalent to the*

inventory position process under an $(0,nQ_{max})$ control policy and compound renewal customer demand.

Explanation. In the $(0,nQ_{max})$ inventory policy with compound renewal customer demand, the inventory position $Y(t)$ decreases at the arrival of a customer order. When the inventory position drops below 0 an amount Q_{max} is ordered. The amount by which the inventory position drops below 0 is called the undershoot and is denoted by U. In the quantity consolidation policy with compound renewal replenishment orders, the collected quantity increases at the arrival of a replenishment order. When the collected quantity exceeds Q_{max}, an amount $\Delta(T) = Q_{max}$ is shipped towards the retail warehouse. In the consolidation process the cumulative order between two shipments is $\Delta(T)$. The inventory position has the same course as the consolidated quantity at a moment in time. The undershoot process U in the $(0,nQ_{max})$ inventory policy is similar to the split order process V in the shipment consolidation process (see Figure 3).

ii) ***Full shipments/flexible truck capacity***

The orders are consolidated until the collected quantity is larger than or equal to a predetermined quantity Q_{max}. The consolidated quantity $\Delta(T)$ is equal to the entire collected quantity in $(0,T]$ and hence $\Delta(T) \geq Q_{max}$. The consolidated quantity $\Delta(T)$ is shipped to the retailer and the consolidation process starts all over again. Figure 4 a) gives a schematic representation of the process. The probability density function of $\Delta(t)$ can be derived from the following proposition.

Proposition 2. *The quantity consolidation process with full shipments/flexible truck capacity under compound renewal demand is equivalent to the inventory position process under an (0,S) control policy and compound renewal customer demand where $S = Q_{max}$.*

Explanation. In the $(0,S)$ inventory policy, where $S = Q_{max}$, with compound renewal customer demand, the inventory position decreases at the arrival of a customer order. When the inventory position drops below 0, an amount $Q_{max} + U$ is ordered such that the inventory position is raised up to Q_{max}. In the quantity consolidation policy with compound renewal replenishment orders, the collected quantity increases at the arrival of a replenishment order. When the collected quantity exceeds Q_{max}, an amount $\Delta(T) = Q_{max} + V$ is shipped towards the retailer. The inventory position has the same course as the consolidated quantity at every moment in time (see Figure 4).

To be able to calculate the waiting time between the placement of an arbitrary order and the departure of the truck, we must derive expressions for the arrival process of orders to be consolidated. To calculate this arrival process, we derive approximations for the order process of the different items from the retailer towards the warehouse. These approximations are described

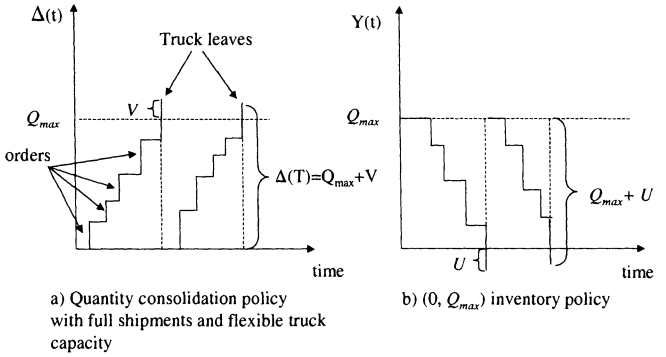


Fig. 4. Schematic representation of the similarity between the shipment consolidation policy with full shipments/flexible truck capacity and the $(0,Q_{max})$ inventory policy

in Appendix A. After that, we superpose the order processes of the different items. The approximations for the superposed process are described in Appendix B. The superposed compound renewal process constitutes the demand for the consolidation process. With this superposed process we can derive expressions for the first two moments of the waiting time due to shipment consolidation. Notice that the superposed process is not a compound renewal process. Yet, our analysis reveals that assuming that the superposed process is a compound renewal process yields good approximations for the performance characteristics. From the first two moments of the waiting time due to shipment consolidation and the first two moments of the transportation time we compute the first two moments of the lead time of an arbitrary order for an item. Using the analysis in Smits et al. (2000) we can compute the reorder levels that yield the required fill rate level and the associated average net stocks.

2.1 Waiting Time Due to Consolidation

In this section, we derive expressions for the waiting time due to shipment consolidation. Again we distinguish between two types of consolidation policies, the time policy and the quantity policy. Due to the compound renewal demand process we cannot hope for exact results of the waiting time distribu-

tion. Our generic approach is to derive expressions for the first two moments of the waiting time and fit a tractable distribution to these first two moments.

Time policy

We assume that the orders arrive at the warehouse according to a compound renewal process. In appendix A and B approximations of the superposed order arrival process are given. This process is independent of the truck departure process, which is a renewal process with deterministic inter-renewal times. Under stationarity it holds that the waiting time until truck departure of an arbitrary order is uniformly distributed on $(0,T)$. The proof of this statement can be found in appendix D. Thus we find

$$E[Z] = \frac{T}{2} \tag{1}$$

$$E[Z^2] = \frac{T^2}{3} \tag{2}$$

Quantity policy

The derivation of the waiting time distribution under the various types of quantity policies is much more complicated than in case of the time policies. Exact results are only available for special cases. Therefore we have to resort to the derivation of approximations for the first two moments of the waiting time distribution.

We have defined $N(X)$ as the number of arrivals between the placement of an arbitrary order and the departure of the truck. For the first moment we obtain

$$E[N(X)] = \sum_{n=1}^{\infty} n(F^{(n-1)*}(Q_{max} - X) - F^{(n)*}(Q_{max} - X)) \tag{3}$$

$$E[N(X)] = \int_0^{Q_{max}} M(Q_{max} - x) dF_X(x) \tag{4}$$

To be able to evaluate this for the different consolidation policies, we have to find an expression for $F_X(x)$ for the two different policies.

i) Partial shipment

We can use Proposition 1 and the fact that the inventory position of an (s,nQ) inventory policy is uniformly distributed between $(s,s+Q)$. Therefore we can conclude that $\Delta(t)$ is uniformly distributed between $(0,Q_{max})$, this gives

$$P\{X \leq x\} = \frac{x}{Q_{max}}.$$

ii) Full shipment/flexible truck capacity

We can use Proposition 2 and the fact that the inventory position of a (s,S) inventory policy is a renewal process. Therefore we can conclude that $\Delta(t)$ is a renewal process, this gives

$$P\{X \leq x\} = \frac{M(x)}{M(Q_{max})}.$$

When O_i is Poisson distributed, we could precisely calculate $N(X)$ for the first two cases, but in practice the coefficient of variation of the replenishment orders are lower than 1. Another difficulty is the calculation of the waiting time from $N(X)$, since we have to take into account that a part of the last replenishment V may not leave directly in the partial shipment case. We observe that it is difficult to find an exact expression for the waiting time. However in practice, it may be possible to standardize the volume of the batchsizes for the different sku's to pallets or boxes and the volume of the truck to a multiple of this volume unit. In this case the consolidated quantity $\Delta(T)$ is exactly equal to the predetermined quantity Q_{max} and we can compute exact derivations for the waiting time.

a. *Equal batchsizes for all items*

In this subsection we derive the first two moments of the waiting time when the volume of the batchsizes of the different items are equal to some Q. The volume of the predetermined shipped quantity Q_{max} is assumed to be a multiple of Q. In this case $\Delta(t)$ has a discrete distribution. The consolidation process starts at the arrival of the first batchsize Q, after this a second batch arrives and the consolidated quantity is $2Q$, then $3Q$ until the predetermined shipped quantity is reached. In the steady state, $\Delta(t)$ is uniformly distributed over $[0, Q, 2Q, ..., (\frac{Q_{max}}{Q} - 1)]$. The different consolidated quantities have the same probability namely $\frac{Q}{Q_{max}}$. It easily follows that

$$E[N(X)] = \sum_{k=0}^{(\frac{Q_{max}}{Q}-1)} k\frac{Q}{Q_{max}} = \frac{1}{2}(\frac{Q_{max}}{Q} - 1)\frac{Q_{max}}{Q}\frac{Q}{Q_{max}} = \frac{1}{2}(\frac{Q_{max}}{Q} - 1) \tag{5}$$

$$E[N(X)^2] = \sum_{k=0}^{(\frac{Q_{max}}{Q}-1)} k^2\frac{Q}{Q_{max}} = \frac{1}{6}(2\frac{Q_{max}^2}{Q^2} - 3\frac{Q_{max}}{Q} + 1) \tag{6}$$

$$E[Z] = E[N(X)]E[R^*] \tag{7}$$

$$E[Z^2] = E[N(X)]\sigma^2(R^*) + E[N(X)^2]E[R^*]^2 \tag{8}$$

b. *Heuristic for non-equal batchsizes*
In the derivations of the waiting time for the quantity policy with non-equal batchsizes we encounter two difficulties: the waiting time is dependent on the quantity consolidation policy (partial shipments or full shipments/flexible truck capacity) and the waiting time may be different for different items. If a batchsize is very large compared to other ones then it is likely that the waiting time for this batchsize is smaller than for the other ones.
We observe that it is not efficient to have a high probability of having more than two batches of the same item in one truck. When there is high probability of having two orders in one truck, we can increase the batchsize without increasing the inventory level which leads to the same inventory costs but may lead to lower transportation and handling costs. To calculate the waiting time of batchsize i, we assume two types of order processes. The order process of item i and the order process of all other items except i. We use the formulas in appendix A and B to calculate the aggregate order process of all products except item i. We define $E[R^*_{\neq i}]$ and $E[R^{*2}_{\neq i}]$ as the first two moment of the inter-arrival times of all other items except item i and we define $E[O^*_{\neq i}]$ and $E[O^{*2}_{\neq i}]$ as the first two moments of the order size of all items except i.

i) Partial shipments
The last order is split such that the consolidated quantity is equal to the predetermined quantity. The part of the last order which is split is denoted by V. Since this consolidation policy is equivalent to the $(0,nQ_{max})$ inventory policy, V is equivalent to the undershoot process in the $(0,nQ_{max})$ inventory model.
For the first two moments of the undershoot process (Appendix B formula's 33 and 34), we use the asymptotic results for the first two moments of the forward recurrence time distribution. Therefore the first two moments of V are

$$E[V] \simeq \frac{E[O^{*2}]}{2E[O^*]} \tag{9}$$

$$E[V^2] \simeq \frac{E[O^{*3}]}{3E[O^*]} \tag{10}$$

Now let us define n_i as the amount of products orders i that are consolidated in an arbitrary consolidated shipment. We neglect the probability that $n_i > 1$, since we deduced previously that it is not cost efficient to have a high probability of having more than 2 orders of the same item in one truck. Figure 5 gives a schematic representation of the truck consolidation process. We define $W_i = O^*_{\neq i} + V + O_i$. Given that $n_i = 1$, the probability that the number of arrivals of order process $\neq i$ is equal to k, is as follows:

$$P\{n_{\neq i} = k | n_i = 1\} = P\{\sum_{i=1}^{k} W_i \leq Q_{max}\} - P\{\sum_{i=1}^{k+1} W_i \leq Q_{max}\} \tag{11}$$

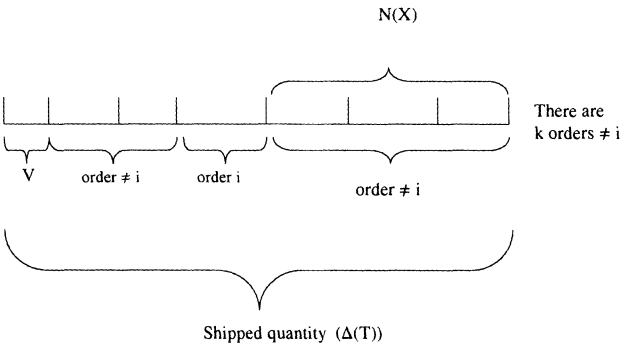


Fig. 5. Schematic representation of the shipment consolidation policy with partial shipments

If O_i is discrete then we can evaluate the formula above by assuming that $\sum_{i=1}^{k} O^*_{\neq i} + V$ is mixed Erlang distributed, else we assume that $\sum_{i=1}^{k} W_i$ is mixed Erlang distributed. Formula (11) gives correct results when k is large and when $O^*_{\neq i}$ and V are exponentially distributed. We assume that the probability that the item i order is the first one is equal to the probability that the item i order is the second one, the third or the last one, given that $n_{\neq i} = k$ and $n_i = 1$. We define $N(X)_i$ as the number of arrivals between the placement of an arbitrary order i and the departure of the truck.

$$E[N(X)_i] \simeq \sum_{k=0}^{\infty} P\{n_{\neq i} = k | n_i = 1\}\left(\frac{1}{k+1}\sum_{s=1}^{k} s + \frac{1}{2}\right) \tag{12}$$

$$E[N(X)_i^2] \simeq \sum_{k=0}^{\infty} P\{n_{\neq i} = k | n_i = 1\}\left(\frac{1}{k+1}\sum_{s=1}^{k} s^2 + \frac{1}{4}\right) \tag{13}$$

The first two moments for the waiting time are

$$E[Z_i] \simeq E[N(X)_i]E[R^*_{\neq i}] \tag{14}$$

$$E[Z_i^2] \simeq E[N(X)_i]\sigma^2(R^*_{\neq i}) + E[N(X)_i^2]E[R^*_{\neq i}]^2 \tag{15}$$

We can derive similar expressions for the full shipment/flexible truck capacity.

3 Simulations and Results

In this section, we test the approximations found for the first two moments of the waiting time. The testing is done with the help of discrete event simulations. The simulations start all with the same seed and stop after 40 000 arrivals of item orders, which ensures accuracy of the simulation results. We assume that the inter-arrival time between customer orders at the retailer is mixed Erlang distributed. Furthermore we assume that the customer order sizes are mixed Erlang distributed.

3.1 Input for the Simulations

For the time policy, we ran 84 different simulations to test the approximations of the waiting time distribution. In the derivations of the aggregate order process some approximations are made to estimate the second moment of the time between two replenishment orders, the approximations perform less if the number of superpositions is smaller than 16. For this reason we varied in the simulations the number of items between the 16 and 32. This is in line with practice where the number of different items is usually larger than 16. All items are identical with $E[A_i] = 1$ and $E[D_i] = 100$, because the derived approximations for the waiting time in time policy are independent of the arrival process of orders to be consolidated. We varied $c^2_{A_i}$ and $c^2_{D_i}$ between 0.4, 1 and 1.6. The batchsizes at the retailer were varied between 500, 1000, 1500, ..., 5000. T is varied between 1, 3 and 6 and P_{2i} between 90 %, 95 % and 99 %. L_d is varied between 2 and 8.

For the quantity policy with equal batchsizes we performed 84 different simulations. The number of different items are varied between 16 and 32. For all items $E[A_i] = 1$ and $c^2_{A_i}$ is varied between 0.2, 1 and 2. In the quantity policy the waiting time due to shipment consolidation is dependent on the arrival process of orders to be consolidated, therefore we assume for the different items different demand processes. We used the fact that in practice, a few items have a large demand and many items have a small demand. We distinguish between two types of items large demand and small demand. 66 % of the items have a small demand. $E[D_i] = 100$ for the large demand and $E[D_i] = 10$ for the small demand. $c^2_{D_i}$ is varied between 0.2, 1 and 2. L_d is varied between 2, 4, 8 and 16. The batchsize is varied between 500 and 2000. Q_{max} is varied between 2000 and 4000 and P_{2i} is varied between 90 %, 95 % and 99 %.

For the quantity policy with non-equal batchsize, we performed 14 different simulations. In the simulations we assumed 16 items with four different types of demand. Type j demand is defined as D_j. The L_d is varied between 2 and 8 and the fill rate is varied between 90 %, 95 % and 99 %.

The input for the cases are given Table 1.

Table 1: Input for the quantity policy simulations with non-equal batch-sizes

Case	$E[A]$	c_A^2	$E[D_1]$	$E[D_2]$	$E[D_3]$	$E[D_4]$	c_D^2	$E[Q]$	Q_{max}
1	1	1	100	150	200	250	1	900	4000
2	1	1	50	70	90	110	1	900	4000
3	1	0.4	50	70	90	110	1	900	4000
4	1	1.6	50	70	90	110	1	900	4000
5	1	1	50	70	90	110	0.4	900	4000
6	1	1	50	70	90	110	1.6	900	4000
7	1	1	50	70	90	110	1	300	8000
8	1	1	50	70	90	110	1	600	8000
9	1	1	50	70	90	110	1	900	8000
10	1	0.4	50	70	90	110	1	900	8000
11	1	1.6	50	70	90	110	1	900	8000
12	1	1	50	70	90	110	0.4	900	8000
13	1	1	50	70	90	110	1.6	900	8000
14	1	1	50	70	90	110	1	2000	8000

where $E[Q]$ is the average batchsize over all different items.

3.2 Results

We used the approximations derived previously in this paper to calculate the first two moments of the waiting time due to shipment consolidation. We fit a mixed Erlang distribution to these first two moments. Using the expressions in Appendix C we computed the reorder levels that ensure the required service levels and the resulting average physical inventory levels. To test our approximations we compare the error in the first two moments of the waiting times, the fill rate and physical inventory level. The error in the first two moments of the waiting time and the average physical inventory level is expressed in a percentage error. The error between the target P_{2i} and the one obtained with the simulation is expressed in absolute percentage error. For every simulation we calculate the relative absolute deviation as follows:

$$RAD_i = \left| P_{2i} - P_{2i}^{target} \right| * 100 \tag{16}$$

The percentage error in the average inventory (PEAI) is calculated as follows:

$$PEAI_i = \frac{\left| E[X_i^+] - E[X_i^{+^*}] \right|}{E[X_i^{+^*}]} * 100 \tag{17}$$

where $E[X_i^{+^*}]$ is the calculated average inventory and $E[X_i^+]$ is the simulated average inventory. To be able to draw meaningful conclusions, we define acceptable margins for the RAD_i values and the $PEAI_i$ values. To construct

a realistic margin, we look at the error in the probability of having backlog $(1 - P_{2i})$. In Table 2 shows good and acceptable values for the fill rate.

Table 2: Good and acceptable fill rates.

P_{2i}^{target}	Good P_{2i}			Acceptable P_{2i}		
	Min	Max	RAD_i values	Min	Max	RAD_i values
90 %	89 %	91 %	1	88 %	92 %	2
95 %	94.5 %	95.5 %	0.5	94 %	96 %	1
99 %	98.9 %	99.1 %	0.1	98.8 %	99.2 %	0.2

For the $PEAI_i$ the good margin is 2.5 % and the acceptable margin is 5%.

Time policy

In this section, we discuss the results of the simulations with time policy. For 84 simulations, the percentage error in the $E[Z]$ is between 0.5 % and 1.5 %, the percentage error in the second moment of the waiting time is between 2 % and 3.5 %. Table 3 summarizes the results.

Table 3: Summary of the results for the "time" policy

P_{2i}	$E[RAD_i]$	RAD_i^m	GP_2	AP_2	$E[PEAI_i]$	$PEAI_i^m$	GI	AI
90 %	0.38	0.93	100 %	100 %	1.03	6.45	83 %	95 %
95 %	0.52	1.17	93 %	98 %	0.79	5.06	90 %	100 %
99 %	0.16	0.36	64 %	78 %	0.57	3.75	90 %	100%

Where RAD_i^m and $PEAI_i^m$ are respectively the maximum RAD and $PEAI$, GP_2 and AP_2 are the percentage within respectively the good and the acceptable range for the error in the fill rate and GI and AI are the percentages within the good and acceptable range for the errors in the average inventory. Hence we conclude that our approach yields excellent results for the time policy.

Quantity policy, equal batchsizes

The results for the situation with batchsizes that are equal in volume are summarized in Table 4.

Table 4: Summary of the results for the quantity policy with equal batchsizes

P_{2i}	$E[RAD_i]$	RAD_i^m	GP_2	AP_2	$E[PEAI_i]$	$PEAI_i^m$	GI	AI
90 %	0.38	1.13	92 %	100 %	1.90	7.46	80 %	99 %
95 %	0.30	0.77	81 %	100 %	1.47	6.53	80 %	99 %
99 %	0.20	0.45	20 %	60 %	1.22	4.32	90 %	99 %

The error in $E[Z]$ is between 0.7 % and 4.2 % and $\sigma^2(Z)$ is between 3.4% and 7.2 %. The peak values are caused by a high probability of having two orders of the same item in one truck. If there is a high probability of having

two orders of the same item in one truck then the demand during the lead time can be no longer approximated by a mixed Erlang distribution because the distribution function of the demand during the lead time has more than one peak. The results were not satisfying in cases of extreme large or low coefficients of variation of the inter-arrival times and order sizes, which is in line with observations about two-moment approximations in general in Tijms (1994). Generally, when the number of items increase, the coefficients of variation will converge to 1 and the results will show better outcome.

Quantity policy, non-equal batchsizes

In section 2.1, we derived approximations for the waiting time due to shipment consolidation for the partial shipment non-equal batchsize quantity policy. In a similar way, we derived expressions for the waiting time due to shipment consolidation for the full shipment/flexible truck capacity policy and for the full shipment/fixed truck capacity. In the full shipment/fixed truck capacity the orders are consolidated until the collected quantity is larger than or equal to a predetermined quantity Q_{max}. In this policy, $\Delta(T) = Q_{max} - O_{N(T)} + V$ is directly shipped to the retailer and the last order $O_{N(T)}$ will be shipped with the next truck to the retailer.

For the three possible consolidation policies the errors in the $E[Z]$ are between 2.58 % and 7.44 % and in $\sigma^2(Z)$ are between 4.90 % and 16.10 %. The results for the fill rate and the average inventory are as follows:

Table 5: Summary of the results for the non-equal batchsizes

Quantity policy	P_{2i}	$E[RAD_i]$	RAD_i^m	GP_2	AP_2	$E[PEAI_i]$	$PEAI_i^m$	GI	AI
1	90 %	0.74	2.22	79 %	97 %	2.16	4.96	64 %	98 %
1	95 %	0.32	1.64	68 %	100 %	2.56	5.46	64 %	98 %
1	99 %	0.15	0.46	46 %	93 %	1.86	4.05	64 %	98 %
2	90 %	0.93	1.99	86 %	97 %	3.22	5.84	52 %	91 %
2	95 %	0.48	1.49	71 %	96 %	2.87	5.29	58 %	91 %
2	99 %	0.15	0.48	61 %	86 %	2.33	5.14	58 %	94 %
3	90 %	1.17	2.20	50 %	86 %	3.65	6.36	0.33	87 %
3	95 %	0.67	1.93	46 %	75 %	3.56	6.02	0.39	87 %
3	99 %	0.19	0.93	43 %	86 %	3.03	6.34	0.33	87 %

In Table 5, quantity policy 1 refers to the partial shipment policy, quantity policy 2 refers to the full shipment /flexible truck capacity policy and quantity policy 3 refers to the full shipment/fixed truck capacity policy.

In quantity policy 2, the fill rates obtained with the simulations where higher than the target fill rates and in quantity policy 3, the fill rates obtained with the simulations where lower than the target fill rates. The errors increase when the coefficients of variation of demand are high (1.6) or low (0.4) due to approximations made in the superposition of mixed Erlang distributions. For example, for the cases 2 to 6, for reference see Table 1, with $P_{2i} = 90\%$

and $L_d = 2$, Table 6 shows the differences in $E[RAD_i]$ for different coefficient of variations.

Table 6: Differences in $E[RAD_i]$ for different coefficient of variations.

c_A^2	1	0.4	1.6	1	1
c_D^2	1	1	1	0.4	1.6
$E[RAD_i]$	0.19	0.60	0.42	0.70	0.74

The errors in the fill rate increase when the average batchsize is large compared to the truck size. Table 7 show this results, the truck size was kept constant at 8000 units and the target fill rate was 90 %.

Table 7: Differences in $E[RAD_i]$ for different number of orders per truck

$\frac{Q_{max}}{E[Q]}$	27	13	9	4
$E[RAD_i]$	0.06	0.1	0.42	1.18

The heuristic performs well as long as the order sizes are not too large compared to the truck size and c_A^2 and c_D^2 are not extremely large or low.

4 Conclusions and Further Research

In this paper we studied the interactions between shipment consolidation policies and inventory management policies. We argued that the lead time of orders from a retailer to a warehouse is influenced by the shipment consolidation policy used at the warehouse. In order to get a deeper insight into this interaction we derived approximations for the waiting time distribution of retailer replenishment orders due to consolidation at the warehouse for different consolidation policies used in practice. An extensive numerical study was conducted to understand and test the different approximations. The study revealed that the approximations performed well. The only problems occurred when the coefficient of variations were very low or very high. Those errors were usually due to approximations made in the superposition of the inter-arrival times. We know from previous research (Smits, de Kok, and van Laarhoven 2000) that if the number of customers and items is large and the demand between customers varies a lot the errors will diminish.

A next step in this research will be to find expressions for the waiting time in a multi-echelon setting and to find close to optimal values for the batchsizes and T in the time policy and close to optimal values for the batchsizes and Q_{max} in the quantity policy. These extensions should enable to develop models for the integrated design of transportation and supply networks that incorporate the operational characteristics of the processes under considerations, such as stochastic demand and stochastic lead times.

APPENDIX

A Replenishment Process

In this section we present a procedure that translates the demand process of an item at the warehouse to a replenishment process of the item towards the central warehouse. These expressions have been derived by Pyke, De Kok and Baganha (1996). Assume O_i is the order process of product i. The first two moments of O_i are derived as follows:

$$E[O_i] = \frac{Q_i E[D_i]}{\int_0^{Q_i} P\{D_i \geq x\} dx} \tag{18}$$

$$E[O_i^2] = Q_i^2 \sum_{z=0}^{\infty} (2z+1) P\{U_i \geq zQ_i\} \tag{19}$$

where

$$P\{U_i \geq zQ_i\} = \frac{\int_0^{Q_i} P\{D_i \geq zQ_i + x\} dx}{\int_0^{Q_i} P\{D_i \geq x\} dx} \tag{20}$$

The time between the placement of two orders is defined as R_i. It is evaluated as follows:

$$R_i = \sum_{j=1} N_i A_{ij} \tag{21}$$

where N_i is defined as the number of arrivals during an arbitrary replenishment cycle at the central warehouse and A_{ij} as the j^{th} inter-arrival time during this replenishment cycle at the central warehouse. There a replenishment cycle is defined as the time that elapses between two consecutive replenishment orders generated by the retailer for product i. Then the first two moments of R_i can be calculated as follows:

$$E[R_i] = E[N_i]E[A_i] \tag{22}$$

$$E[R_i^2] = E[N_i]\sigma^2(A_i) + E[N_i^2]E^2[A_i] \tag{23}$$

Due to flow conservation the following relation holds for $E[N_i]$:

$$E[N_i] = \frac{E[O_i]}{E[D_i]} \tag{24}$$

To determine an expression for $E[N_i^2]$, we apply the following approximation which is accurate when $\frac{Q_i}{E[D_i]}$ is not too small ($\frac{Q_i}{E[D_i]} > 1$):

$$E[N_i^2] \simeq \left(\frac{Q_i^2}{E^2[D_i]} + c_{D_i}^2 \frac{Q_i}{E[D_i]} + \frac{E^2[D_i^2]}{2E^4[D_i]} - \frac{E[D_i^3]}{3E^3[D_i]} \right) \frac{E[O_i]}{Q_i} \tag{25}$$

B Aggregate Order Process

In this paragraph we will find expressions for the first two moments of the inter-arrival time and the order size of an arbitrary order towards the central warehouse. To do this, we apply the stationary interval method developed by Whitt (1982), to superpose renewal processes. Instead of superposing hyper-exponential and shifted exponential distributions we superpose mixtures of Erlang distributions. In the superposition procedure it is assumed that the superposed process is a renewal process, which is not true. Because when we superpose N renewal processes, the first renewal time of the superposed process should be the minimum of the first renewal times of the N individual renewal processes. The superposition gives exact results when the renewal processes are Poisson distributed. The superposed process converges to the Poisson process when N tends to infinity (Tijms (1994)). The N products are represented by index i. The first two moments of T^* and O^* are calculated as follows: (see De Kok (1996))

$$E[R^*] = \frac{1}{\sum_{i=1}^{N} \frac{1}{E[R_i]}} \tag{26}$$

$$E[R^{*2}] \simeq 2E[R^*] \int_0^\infty \left(\prod_{i=1}^{N} \frac{1}{E[R_i]} \right) \left(\prod_{i=1}^{N} \int_x^\infty (1 - F_{R_i}(y)) dy \right) dx \tag{27}$$

$$E[O^*] = \sum_{i=1}^{N} \frac{E[R^*]}{E[R_i]} E[O_i] \tag{28}$$

$$E[O^{*2}] = \sum_{i=1}^{N} \frac{E[R^*]}{E[R_i]} E[O_i^2] \tag{29}$$

C Derivation of the Reorder Level and Physical Inventory Level

First, we derive some analytical approximations to calculate the reorder levels for a target fill rate. Given the inter-arrival time A_i of each item and its

demand size D_i, the reorder levels can be analytically evaluated. The reorder levels s_i are calculated as follows: (see Janssen (1998) for a proof)

$$P_{2i} \simeq 1 - \frac{E[(D_i(L_i^*) + U_i - s_i)^+] - E[(D_i(L_i^*) + U_i - s_i - Q_i)^+]}{Q_i} \tag{30}$$

It is possible to evaluate the reorder level s_i using the bisection method. L^* is the total lead time, it is expressed as the sum of the waiting time due to consolidation and the transportation time. $D_i(L_i^*)$ is the demand for product i at the warehouse during the lead time L_i^*. The mean and the variance of $D_i(L_i^*)$ are calculated as follows: (For a detailed explanation see De Kok (1991))

$$E[D_i(L_i^*)] \simeq (\frac{E[L_i^*]}{E[A]} + \frac{E[A^2]}{2E^2[A]} - 1)E[D_i] \tag{31}$$

$$\begin{aligned}\sigma^2(D_i(L_i^*)) \simeq & \frac{E[L_i^*]}{E[A]}\sigma^2(D_i) + \frac{E[L_i^*]}{E[A]}c_A^2 E^2[D_i] + \sigma^2(L_i^*)\frac{E^2[D_i]}{E^2[A]} \\ & + \frac{(c_A^2 - 1)}{2}\sigma^2(D_i) + \frac{(1 - c_A^4)}{12}E^2[D_i]\end{aligned} \tag{32}$$

Expressions for the first and the second moment of the undershoot are

$$E[U_i] \simeq \frac{E[D_i^2]}{2E[D_i]} \tag{33}$$

$$E[U_i^2] \simeq \frac{E[D_i^3]}{3E[D_i]} \tag{34}$$

For a derivation of these results see Tijms (1994). The average stock on hand is calculated as follows:

$$E[X_i^+] \simeq \frac{1}{2Q_i}(E[(s_i + Q_i - D_i(L_i^*))^{+^2}] - E[(s_i - D_i(L_i^*))^{+^2}]) \tag{35}$$

For a derivation see Janssen (1998) or De Kok (1991).

D Derivation of Z for the Time Policy

Theorem 3. *Z is uniformly distributed over the interval (0,T) for time policy.*

Proof. We define $\tilde{A}$ as the residual life time of the inter-arrival time of an arbitrary customer at an arbitrary moment in time. We define W as the residual lifetime of the truck arrival process at the arbitrary moment in time. Since W is the time between an arbitrary moment in time and the departure of the truck, W is uniformly distributed over $(0,T)$, for references see Doob (1953). $k \in N$

$$\tilde{A} + Z = W + kT$$

$$\tilde{A} + Z + T - W = (k+1)T$$

We define $X = T - W$, it is easy to see that X is uniform distributed between $(0,T)$ and $\tilde{k} = k + 1$

$$\begin{aligned}
P\{Z \le z\} &= P\{X + \tilde{A} \in (\tilde{k}T - z, \tilde{k}T), \tilde{k} \in N\} \\
&= \sum_{\tilde{k}=1}^{\infty} \tfrac{1}{T} \int_0^T P\{L_s \in (\tilde{k}T - z - x, \tilde{k}T - x)\} dx \\
&= \sum_{\tilde{k}=1}^{\infty} \tfrac{1}{T} \int_0^T [\int_{\tilde{k}T-z-x}^{\infty} dF_{\tilde{A}}(\tilde{a}) - \int_{\tilde{k}T-x}^{\infty} dF_{\tilde{A}}(\tilde{a})] dx \\
&= \tfrac{1}{T}[\int_0^{T-z} \int_{T-z-x}^{\infty} dF_{\tilde{A}}(\tilde{a}) dx + \int_{T-z}^{0} \int_0^{\infty} dF_{\tilde{A}}(\tilde{a}) dx - \int_0^T \int_{T-x}^{\infty} dF_{\tilde{A}}(\tilde{a}) dx] \\
&\quad + \sum_{\tilde{k}=2}^{\infty} \tfrac{1}{T}[\int_{\tilde{k}T-z}^{\infty} \int_0^T dx dF_{\tilde{A}}(\tilde{a}) + \int_{((\tilde{k}-1)T-z)}^{\tilde{k}T-z} \int_{(\tilde{k}T-z-\tilde{a})}^{T} dx dF_{\tilde{A}}(\tilde{a}) \\
&\quad - \int_{\tilde{k}T}^{\infty} \int_0^T dx dF_{\tilde{A}}(\tilde{a}) - \int_{((\tilde{k}-1)T)}^{\tilde{k}T} \int_{(\tilde{k}T-\tilde{a})}^{T} dx dF_{\tilde{A}}(\tilde{a})] \\
&= \tfrac{1}{T}[\int_0^{T-z} \int_{T-z-\tilde{a}}^{T-z} dx dF_{\tilde{A}}(\tilde{a}) + \int_{T-z}^{\infty} \int_0^{T-z} dx dF_{\tilde{A}}(\tilde{a}) - \int_T^{\infty} \int_0^T dx dF_{\tilde{A}}(\tilde{a}) \\
&\quad + z - \int_0^T \int_{T-\tilde{a}}^{T} dx dF_{\tilde{A}}(\tilde{a})] + \sum_{\tilde{k}=2}^{\infty} [\int_{(\tilde{k}T-z)}^{\infty} dF_{\tilde{A}}(\tilde{a}) - \int_{\tilde{k}T}^{\infty} dF_{\tilde{A}}(\tilde{a}) \\
&\quad + \int_{(\tilde{k}-1)T-z}^{\tilde{k}T-z} (\tfrac{\tilde{a}-z-(\tilde{k}-1)T}{T}) dF_{\tilde{A}}(\tilde{a}) - \int_{(\tilde{k}-1)T}^{\tilde{k}T} \tfrac{\tilde{a}-(\tilde{k}-1)T}{T} dF_{\tilde{A}}(\tilde{a})] \\
&= \tfrac{z}{T} + \int_{T-z}^{\infty} \tfrac{T-z}{T} dF_{\tilde{A}}(\tilde{a}) - \int_{T-z}^{T} \tfrac{\tilde{a}}{T} dF_{\tilde{A}}(\tilde{a}) - \int_T^{\infty} dF_{\tilde{A}}(\tilde{a}) \\
&\quad + \sum_{\tilde{k}=2}^{\infty} [\int_{(\tilde{k}-1)T-z}^{\tilde{k}T-z} \tfrac{z}{T} dF_{\tilde{A}}(\tilde{a}) + \int_{(\tilde{k}-1)T-z}^{(\tilde{k}-1)T} \tfrac{\tilde{a}-(\tilde{k}-1)T}{T} dF_{\tilde{A}}(\tilde{a}) \\
&\quad - \int_{\tilde{k}T-z}^{\tilde{k}T} \tfrac{\tilde{a}-\tilde{k}T}{T} dF_{\tilde{A}}(\tilde{a})] \\
&= \tfrac{z}{T} + \int_{T-z}^{\infty} \tfrac{T-z}{T} dF_{\tilde{A}}(\tilde{a}) - \int_{T-z}^{T} \tfrac{\tilde{a}}{T} dF_{\tilde{A}}(\tilde{a}) - \int_T^{\infty} dF_{\tilde{A}}(\tilde{a}) \\
&\quad + \int_{T-z}^{\infty} \tfrac{z}{T} dF_{\tilde{A}}(\tilde{a}) + \int_{T-z}^{T} \tfrac{\tilde{a}-T}{T} dF_{\tilde{A}}(\tilde{a}) \\
&= \tfrac{z}{T}
\end{aligned}$$

References

1 **Doob, J.L. (1953)***Stochastic processes* . Wiley London

2 **Federgruen, A., H. Groenevelt and H.C. Tijms (1984)** *Coordinated replenishments in a multi-item inventory system with compound Poisson demands* . Management Science 30, p 344-357

3 **Goyal, S.K. and A.T. Satir, (1989)** *Joint replenishment inventory control: deterministic and stochastic models* . European Journal of Operational research 38, p 3-13

4 **Higginson, J.K. and J.H. Bookbinder, (1994)** *Policy recommendations for a shipment consolidation program.* Journal of Business Logistics 15, p 87-112

5 **Higginson, J.K. and J.H. Bookbinder, (1995a)** *Probabilistic models of freight consolidation systems.* Working paper, Department of Management Science, University of Waterloo, 1995

6 **Higginson, J.K. and J.H. Bookbinder, (1995b)** *Markovian Decision Processes in Shipment Consolidation.* Transportation Science 29, nb. 3

7 **Janssen, F.B.S.L.P. (1998)** *Inventory management systems.* Ph.D. thesis, Center for economic research, Tilburg University, The Netherlands

8 **Kok, A.G. de (1991)***Basics of Inventory Management.* FEW working papers 520-525, Tilburg University, The Netherlands

9 **Kok, A.G. de (1996)** *Analysis of (s,nQ)-installation stock policies in divergent multi-echelon networks.* Research report TUE/TM/LBS/96-09, Eindhoven University of Technology, The Netherlands.

10 **Liu, L. and X. Yuan, (2000)** *Coordinating replenishments in inventory systems with correlated demands.* European Journal of Operations Research 123, p 490-503

11 **Pyke, D.A., A.G. de Kok and M. Baganha (1996)** *The undershoot distribution in (s,nQ)-models* . Working paper TUE/TM/LBS/96-05, Eindhoven University of Technology, The Netherlands.

12 **Smits, S.R., A.G. De Kok and P.J.M. Van Laarhoven (2000)** *Analysis of Divergent N-echelon (s,nQ)-policies Under Compound Renewal Demand.* Submitted for publication

13 **Tijms, H.C. (1994)** *Stochastic models: an algorithmic approach.* John Wiley & Sons.

14 **Viswanathan, V., (2000)** *Periodic review (s,S) policies for joint replenishment inventory systems* . Working paper, Nanyang Business School, Nanyang tech. University, Singapore 1996

15 **Whitt, W. (1982)** *Approximating a point process by a renewal process, I: Two basic methods* . Operations Research 30, p 125-147.

List of Contributors

E. **Angelelli**, Department of Quantitative Methods, University of Brescia, Contrada S. Chiara, 46/b, 25122 Brescia, Italy

A. **Bauer**, Fraunhofer Anwendungszentrum für Verkehrslogistik und Kommunikationstechnik, 90489 Nürnberg, Germany

J. D. **Blackburn**, Owen Graduate School of Management, Vanderbilt University, Nashville, Tennessee, TN 37203, USA

J. M. **Bloemhof-Ruwaard**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

M. **Bjørndal**, The Norwegian School of Economics and Business Administration, Department of Finance and Management Science, Helleveien 30, 5045 Bergen, Norway

P. **Chevalier**, Institut d'Administration et de Gestion, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

J. R. **Daduna**, University of Applied Business Administration, Badensche Straße 50–51, 10825 Berlin, Germany

R. **Dillmann**, Department of Economics and Business Administration, University of Wuppertal, 42097 Wuppertal, Germany

A. **Drexl**, Christian-Albrechts-Universität zu Kiel, Lehrstuhl für Produktion und Logistik, 24908 Kiel, Germany

K. **Fjell**, The Norwegian School of Economics and Business Administration, Helleveien 30, 5045 Bergen, Norway

C. **Gotzel**, Faculty of Economics and Management, Otto-von-Guericke University of Magdeburg, 39016 Magdeburg, Germany

V. D. R. **Guide, Jr.**, Duquesne University, 600 Forbes Ave., Pittsburgh, PA 15241, USA

K. **Inderfurth**, Faculty of Economics and Management, Otto-von-Guericke University of Magdeburg, 39016 Magdeburg, Germany

K. **Jørnsten**, The Norwegian School of Economics and Business Administration, Department of Finance and Management Science, Helleveien 30, 5045 Bergen, Norway

A. **Klose**, University of St. Gallen, 9000 St. Gallen, Switzerland

A. G. **de Kok**, Faculty of Technology Management, Technische Universiteit Eindhoven, Pav E6, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

R. **de Koster**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

A. **Kraal**, Logistics Center of Expertise, Campina Melkunie B.V., Woerden, The Netherlands

H. **Krikke**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

E. **van der Laan**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

A. **van der Linden**, Logistics Center of Expertise, Campina Melkunie B.V., Woerden, The Netherlands

R. **Mansini**, Department of Electronics for Automation, University of Brescia, via Branze 38, 25133 Brescia, Italy

M. **Mazzarino**, Dipartimento di Pianificazione, IUAV – Istituto Universitario di Architettura di Venezia, *c/o* ISTIEE – Università degli Studi di Trieste

J. R. **van der Meer**, IG&H Management Consultants, Woerden, The Netherlands

J. A. E. E. **van Nunen**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

C. P. **Pappis**, Department of Industrial Management, University of Piraeus, 80, Karaoli & Dimitriou Str. 18534 Piraeus, Greece

R. **Pesenti**, Dipartimento di Ingegneria Automatica ed Informatica (DIAI), Università degli Studi di Palermo

Y. **Pochet**, CORE, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

H. **E. Romeijn**, Department of Industrial and Systems Engineering, University of Florida, 303 Weil Hall, P.O. Box 116595, Gainesville, Florida 32611-6595

D. **Romero Morales**, Faculty of Economics and Business Administration, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

K. J. **Roodbergen**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

S. R. **Smits**, Faculty of Technology Management, Technische Universiteit Eindhoven, Pav E6, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

M. G. **Speranza**, Department of Quantitative Methods, University of Brescia, Contrada S. Chiara, 46/b, 25122 Brescia, Italy

L. **Talbot**, Institut d'Administration et de Gestion, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

R. **Teunter**, Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

G. T. **Tsoulfas**, Department of Industrial Management, University of Piraeus, 80, Karaoli & Dimitriou Str. 18534 Piraeus, Greece

W. **Ukovich**, Dipartimento di Elettrotecnica, Elettronica ed Informatica (DEEI), Università degli Studi di Trieste

I. F. A. **Vis**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

J. **Vroom**, Logistics Center of Expertise, Campina Melkunie B.V., Woerden, The Netherlands

M. **Wagner**, Department for Production and Logistics, University of Augsburg, Universitätsstraße 16, 86135 Augsburg, Germany

L. N. **Van Wassenhove**, INSEAD, Boulevard de Constance, 77305 Fontainebleau Cedex, France

Lecture Notes in Economics and Mathematical Systems

For information about Vols. 1–429
please contact your bookseller or Springer-Verlag

Vol. 430: J. R. Daduna, I. Branco, J. M. Pinto Paixão (Eds.), Computer-Aided Transit Scheduling. XIV, 374 pages. 1995.

Vol. 431: A. Aulin, Causal and Stochastic Elements in Business Cycles. XI, 116 pages. 1996.

Vol. 432: M. Tamiz (Ed.), Multi-Objective Programming and Goal Programming. VI, 359 pages. 1996.

Vol. 433: J. Menon, Exchange Rates and Prices. XIV, 313 pages. 1996.

Vol. 434: M. W. J. Blok, Dynamic Models of the Firm. VII, 193 pages. 1996.

Vol. 435: L. Chen, Interest Rate Dynamics, Derivatives Pricing, and Risk Management. XII, 149 pages. 1996.

Vol. 436: M. Klemisch-Ahlert, Bargaining in Economic and Ethical Environments. IX, 155 pages. 1996.

Vol. 437: C. Jordan, Batching and Scheduling. IX, 178 pages. 1996.

Vol. 438: A. Villar, General Equilibrium with Increasing Returns. XIII, 164 pages. 1996.

Vol. 439: M. Zenner, Learning to Become Rational. VII, 201 pages. 1996.

Vol. 440: W. Ryll, Litigation and Settlement in a Game with Incomplete Information. VIII, 174 pages. 1996.

Vol. 441: H. Dawid, Adaptive Learning by Genetic Algorithms. IX, 166 pages.1996.

Vol. 442: L. Corchón, Theories of Imperfectly Competitive Markets. XIII, 163 pages. 1996.

Vol. 443: G. Lang, On Overlapping Generations Models with Productive Capital. X, 98 pages. 1996.

Vol. 444: S. Jørgensen, G. Zaccour (Eds.), Dynamic Competitive Analysis in Marketing. X, 285 pages. 1996.

Vol. 445: A. H. Christer, S. Osaki, L. C. Thomas (Eds.), Stochastic Modelling in Innovative Manufactoring. X, 361 pages. 1997.

Vol. 446: G. Dhaene, Encompassing. X. 160 pages. 1997.

Vol. 447: A. Artale, Rings in Auctions. X. 172 pages. 1997.

Vol. 448: G. Fandel, T. Gal (Eds.), Multiple Criteria Decision Making. XII, 678 pages. 1997.

Vol. 449: F. Fang, M. Sanglier (Eds.), Complexity and Self-Organization in Social and Economic Systems. IX, 317 pages, 1997.

Vol. 450: P. M. Pardalos, D. W. Hearn, W. W. Hager, (Eds.), Network Optimization. VIII, 485 pages, 1997.

Vol. 451: M. Salge, Rational Bubbles. Theoretical Basis, Economic Relevance, and Empirical Evidence with a Special Emphasis on the German Stock Market. IX, 265 pages. 1997.

Vol. 452: P. Gritzmann, R. Horst, E. Sachs, R. Tichatschke (Eds.), Recent Advances in Optimization. VIII, 379 pages. 1997.

Vol. 453: A. S. Tangian, J. Gruber (Eds.), Constructing Scalar-Valued Objective Functions. VIII, 298 pages. 1997.

Vol. 454: H.-M. Krolzig, Markov-Switching Vector Autoregressions. XIV, 358 pages. 1997.

Vol. 455: R. Caballero, F. Ruiz, R. E. Steuer (Eds.), Advances in Multiple Objective and Goal Programming. VIII, 391 pages. 1997.

Vol. 456: R. Conte, R. Hegselmann, P. Terna (Eds.), Simulating Social Phenomena. VIII, 536 pages. 1997.

Vol. 457: C. Hsu, Volume and the Nonlinear Dynamics of Stock Returns. VIII, 133 pages. 1998.

Vol. 458: K. Marti, P. Kall (Eds.), Stochastic Programming Methods and Technical Applications. X, 437 pages. 1998.

Vol. 459: H. K. Ryu, D. J. Slottje, Measuring Trends in U.S. Income Inequality. XI, 195 pages. 1998.

Vol. 460: B. Fleischmann, J. A. E. E. van Nunen, M. G. Speranza, P. Stähly, Advances in Distribution Logistic. XI, 535 pages. 1998.

Vol. 461: U. Schmidt, Axiomatic Utility Theory under Risk. XV, 201 pages. 1998.

Vol. 462: L. von Auer, Dynamic Preferences, Choice Mechanisms, and Welfare. XII. 226 pages. 1998.

Vol. 463: G. Abraham-Frois (Ed.), Non-Linear Dynamics and Endogenous Cycles. VI, 204 pages. 1998.

Vol. 464: A. Aulin, The Impact of Science on Economic Growth and its Cycles. IX, 204 pages. 1998.

Vol. 465: T. J. Stewart, R. C. van den Honert (Eds.), Trends in Multicriteria Decision Making. X. 448 pages. 1998.

Vol. 466: A. Sadrieh, The Alternating Double Auction Market. VII, 350 pages. 1998.

Vol. 467: H. Hennig-Schmidt, Bargaining in a Video Experiment. Determinants of Boundedly Rational Behavior. XII, 221 pages. 1999.

Vol. 468: A. Ziegler, A Game Theory Analysis of Options. XIV, 145 pages. 1999.

Vol. 469: M. P. Vogel, Environmental Kuznets Curves. XIII, 197 pages. 1999.

Vol. 470: M. Ammann, Pricing Derivative Credit Risk. XII, 228 pages. 1999.

Vol. 471: N. H. M. Wilson (Ed.), Computer-Aided Transit Scheduling. XI. 444 pages. 1999.

Vol. 472: J.-R. Tyran, Money Illusion and Strategic Complementarity as Causes of Monetary Non-Neutrality. X, 228 pages. 1999.

Vol. 473: S. Helber, Performance Analysis of Flow Lines with Non-Linear Flow of Material. IX. 280 pages. 1999.

Vol. 474: U. Schwalbe, The Core of Economies with Asymmetric Information. IX, 141 pages. 1999.

Vol. 475: L. Kaas, Dynamic Macroeconomics with Imperfect Competition. XI, 155 pages. 1999.

Vol. 476: R. Demel, Fiscal Policy, Public Debt and the Term Structure of Interest Rates. X, 279 pages. 1999.

Vol. 477: M. Théra, R. Tichatschke (Eds.), Ill-posed Variational Problems and Regularization Techniques. VIII, 274 pages. 1999.

Vol. 478: S. Hartmann, Project Scheduling under Limited Resources. XII, 221 pages. 1999.

Vol. 479: L. v. Thadden, Money, Inflation, and Capital Formation. IX, 192 pages. 1999.

Vol. 480: M. Grazia Speranza, P. Stähly (Eds.), New Trends in Distribution Logistics. X, 336 pages. 1999.

Vol. 481: V. H. Nguyen, J. J. Strodiot, P. Tossings (Eds.), Optimation. IX, 498 pages. 2000.

Vol. 482: W. B. Zhang, A Theory of International Trade. XI, 192 pages. 2000.

Vol. 483: M. Königstein, Equity, Efficiency and Evolutionary Stability in Bargaining Games with Joint Production. XII, 197 pages. 2000.

Vol. 484: D. D. Gatti, M. Gallegati, A. Kirman, Interaction and Market Structure. VI, 298 pages. 2000.

Vol. 485: A. Garnaev, Search Games and Other Applications of Game Theory. VIII, 145 pages. 2000.

Vol. 486: M. Neugart, Nonlinear Labor Market Dynamics. X, 175 pages. 2000.

Vol. 487: Y. Y. Haimes, R. E. Steuer (Eds.), Research and Practice in Multiple Criteria Decision Making. XVII, 553 pages. 2000.

Vol. 488: B. Schmolck, Ommitted Variable Tests and Dynamic Specification. X, 144 pages. 2000.

Vol. 489: T. Steger, Transitional Dynamics and Economic Growth in Developing Countries. VIII, 151 pages. 2000.

Vol. 490: S. Minner, Strategic Safety Stocks in Supply Chains. XI, 214 pages. 2000.

Vol. 491: M. Ehrgott, Multicriteria Optimization. VIII, 242 pages. 2000.

Vol. 492: T. Phan Huy, Constraint Propagation in Flexible Manufacturing. IX, 258 pages. 2000.

Vol. 493: J. Zhu, Modular Pricing of Options. X, 170 pages. 2000.

Vol. 494: D. Franzen, Design of Master Agreements for OTC Derivatives. VIII, 175 pages. 2001.

Vol. 495: I Konnov, Combined Relaxation Methods for Variational Inequalities. XI, 181 pages. 2001.

Vol. 496: P. Weiß, Unemployment in Open Economies. XII, 226 pages. 2001.

Vol. 497: J. Inkmann, Conditional Moment Estimation of Nonlinear Equation Systems. VIII, 214 pages. 2001.

Vol. 498: M. Reutter, A Macroeconomic Model of West German Unemployment. X, 125 pages. 2001.

Vol. 499: A. Casajus, Focal Points in Framed Games. XI, 131 pages. 2001.

Vol. 500: F. Nardini, Technical Progress and Economic Growth. XVII, 191 pages. 2001.

Vol. 501: M. Fleischmann, Quantitative Models for Reverse Logistics. XI, 181 pages. 2001.

Vol. 502: N. Hadjisavvas, J. E. Martínez-Legaz, J.-P. Penot (Eds.), Generalized Convexity and Generalized Monotonicity. IX, 410 pages. 2001.

Vol. 503: A. Kirman, J.-B. Zimmermann (Eds.), Economics with Heterogenous Interacting Agents. VII, 343 pages. 2001.

Vol. 504: P.-Y. Moix (Ed.),The Measurement of Market Risk. XI, 272 pages. 2001.

Vol. 505: S. Voß, J. R. Daduna (Eds.), Computer-Aided Scheduling of Public Transport. XI, 466 pages. 2001.

Vol. 506: B. P. Kellerhals, Financial Pricing Models in Continuous Time and Kalman Filtering. XIV, 247 pages. 2001.

Vol. 507: M. Koksalan, S. Zionts, Multiple Criteria Decision Making in the New Millenium. XII, 481 pages. 2001.

Vol. 508: K. Neumann, C. Schwindt, J. Zimmermann, Project Scheduling with Time Windows and Scarce Resources. XI, 335 pages. 2002.

Vol. 509: D. Hornung, Investment, R&D, and Long-Run Growth. XVI, 194 pages. 2002.

Vol. 510: A. S. Tangian, Constructing and Applying Objective Functions. XII. 582 pages. 2002.

Vol. 511: M. Külpmann, Stock Market Overreaction and Fundamental Valuation. IX, 198 pages. 2002.

Vol. 512: W.-B. Zhang, An Economic Theory of Cities.XI, 220 pages. 2002.

Vol. 513: K. Marti, Stochastic Optimization Techniques. VIII, 364 pages. 2002.

Vol. 514: S. Wang, Y. Xia, Portfolio and Asset Pricing. XII, 200 pages. 2002.

Vol. 515: G. Heisig, Planning Stability in Material Requirements Planning System. XII, 264 pages. 2002.

Vol. 516: B. Schmid, Pricing Credit Linked Financial Instruments. X, 246 pages. 2002.

Vol. 517: H. I. Meinhardt, Cooperative Decision Making in Common Pool Situations. VIII, 205 pages. 2002.

Vol. 518: S. Napel, Bilateral Bargaining. VIII, 188 pages. 2002.

Vol. 519: A. Klose, G. Speranza, L. N. Van Wassenhove (Eds.), Quantitative Approaches to Distribution Logistics and Supply Chain Management. XIII, 421 pages. 2002.