Lecture Notes in Economics and Mathematical Systems 544

B. Fleischmann
A. Klose (Eds.)

Distribution Logistics

Advanced Solutions to Practical Problems

Springer

Editors

Prof. Dr. Bernhard Fleischmann
University of Augsburg
Department of Production and Logistics
Universitätsstraße 16
86159 Augsburg, Germany

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

Library of Congress Control Number: 2004107326

ISSN 0075-8442
ISBN 978-3-540-22100-5 ISBN 978-3-642-17020-1 (eBook)
DOI 10.1007/978-3-642-17020-1

springeronline.com

Originally published by Springer-Verlag Berlin Heidelberg New York in 2005

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

Printed on acid-free paper 42/3130Di 5 4 3 2 1 0

Editorial

This book is the fourth volume on applied research in Distribution Logistics that results from the work of a group of mainly European researchers meeting regularly at the "IWDL" (International Workshop on Distribution Logistics). The book contains a selection of research papers, some of which have been presented at the IWDL 7 in Grainau, Germany, in October 2002. It continues the tradition of the previous volumes, all appeared as Springer Lecture Notes in Economics and Mathematical Systems (no. 460, 1998; no. 480, 1999; no. 519, 2002), which have found a favourable acceptance by the logistics community. Recently, the second volume has appeared in a Chinese translation.

Distribution Logistics make up a major part of Supply Chain Management and concern all flows of goods and information between the production sites and the customers. Various trends contribute to a considerable increase in complexity of distribution systems, such as the globalization of the business of most manufacturers and the increasing dynamics of the customer demands. It is therefore not surprising, that the interest in "advanced" planning methods, based on quantitative optimization, increases in practice, in particular for the design of distribution systems and the control of the various interrelated transportation and warehousing processes. This development is favoured by the advances in the information and communication technology, which has enabled a nearly unlimited availability of data at any place at any time. On this background, a new kind of supply chain planning software for manufacturing industries, the "Advanced Planning Systems", have emerged. Their usefulness for Distribution Logistics however is restricted, as the manufacturers have outsourced major tasks of it to logistics service providers (LSP), who combine the distribution processes for several supply chains. Therefore, Distribution Logistics is mainly the responsibility of the LSP. But Advanced Planning Systems for an LSP do not exist yet. The papers in this book deal with some of these developments and challenges.

This book, like its predecessors, includes papers on a recent branch of Logistics, Reverse Logistics. The reason for this is threefold: First, Reverse Logistics are closely related to Distribution Logistics, as the reverse flows link again customers and production sites. Second, the practical importance of Reverse Logistics is fast-growing, and third, a number of members in the IWDL group have become experts in this field.

The 13 papers of this volume have been arranged in four chapters, the first three on the traditional subjects: design of distribution systems, tactical and operational vehicle routing and warehousing operations. The last chapter is dedicated to Reverse Logistics.

Chapter 1 addresses theory and application of the distribution network design and distribution concepts for E-commerce. *Görtz and Klose* consider the general Capacitated Facility Location Problem (CFLP), which is a basic model for various network design problems. They review the existing solution methods for this hard combinatorial problem and suggest a new type of lower bounds based on column generation. The lower bound is used in a branch & bound algorithm and tested in a computational experiment. *Bauer* stresses the need for considering cost and capacities of warehouse processes when designing a distribution network. She introduces a "Modular Node Model" for that purpose and shows the impact of the warehouse processes on the resulting network. *Bloemhof, Smeets and van Nunen* report an interesting practical case of supply chain design for the Dutch pig husbandry. The network consisting of farmers, slaughter houses, wholesalers and retailers is optimized using a mixed integer programming model. *Daduna and Lenz* deal with the often neglected aspects of physical distribution caused by "Online-Shopping". They investigate its impact on the traffic for the "last mile", both for the commercial freight deliveries and the private shopping trips.

Chapter 2 deals with different cases of vehicle routing in various practical environments. *Mansini, Speranza and Angelelli* address on-line routing problems occurring in the multi-depot network of an LSP. They analyse the particular planning situation and present a model and new algorithms. *Archetti and Speranza* report on the solution of a case of waste collection in the county of Brescia, Italy. It contains a pickup and delivery problem for the transport of empty and full containers, where various side constraints have to be considered. *Bieding* deals with planning time-critical standard-tours, as they occur in the daily distribution of newspapers. A focus is on the improvement of data on travel times using modern measuring systems such as GPS and RFID. *Schönberger and Kopfer* consider a general pickup and delivery problem, where a distinction is made between own vehicles, which are to be routed, and subcontracted vehicles, which are paid per load.

The papers in **Chapter 3**, again, concern routing problems, but for order-picking within a warehouse. *Le Anh and de Koster* investigate the performance of various dispatching rules for the on-line control of the vehicles. They report a simulation study in two real-life environments. *Le Duc and de Koster* present models for estimating the length of an order-picking tour, depending on the storage strategy. These approximations allow the optimization of the boundaries between the warehouse zones and of the warehouse layout.

Finally, **Chapter 4** consists of contributions to the field of Reverse Logistics. For the case of reusing components of products after their end of live, *Geyer and Van Wassenhove* study the effect of time constraints, such as a limited component durability or a finite product life cycle. In the same context, *Krikke, van Nunen, Zuidwijk and Kuik* consider the management of return flows, using information on the "installed base" of product placements in the market. *De Brito, Dekker and Flapper* provide a comprehensive review of case studies in Reverse Logistics, classifying the content of more than 60 publications.

The editors are indebted to all authors for their valuable contributions and to the referees whose work was essential to ensure a high quality level of this book.

Bernhard Fleischmann, University of Augsburg, Germany
Andreas Klose, University of Zurich, Switzerland

July 2004

Contents

Chapter 1: Networks

Chapter 2: Vehicle Routing

Chapter 3: Warehousing

Chapter 4: Reverse Logistics

Appendix

Chapter 1

Networks

An Exact Column Generation Approach to the Capacitated Facility Location Problem

Andreas Klose[1] and Simon Görtz[2]

[1] University of St. Gallen and Institute for Operations Research, University of Zurich, Switzerland
[2] Faculty of Economics and Social Sciences, University of Wuppertal, Germany

Abstract The Capacitated Facility Location Problem (CFLP) is a well-known combinatorial optimization problem with applications in distribution and production planning. It consists in selecting plant sites from a finite set of potential sites and in allocating customer demands in such a way as to minimize operating and transportation costs. A variety of lower bounds based on Lagrangean relaxation and subgradient optimization has been proposed for this problem. However, in order to solve large or difficult problem instances information about a primal (fractional) solution is important. Therefore, we employ column generation in order to solve a corresponding master problem exactly. The algorithm uses different strategies for stabilizing the column generation process. Furthermore, the column generation method is employed within a branch-and-price procedure for computing optimal solutions to the CFLP. Computational results are reported for a set of larger and difficult problem instances. The results are compared with computational results obtained from a branch-and-bound procedure based on Lagrangean relaxation and subgradient optimization and a branch-and-bound method that uses the LP relaxation and polyhedral cuts.

1 Introduction

The Capacitated Facility Location Problem (CFLP) is a well-known and well studied combinatorial optimization problem. It consists in deciding which plants to open from a given set of potential plant locations and how to assign customers to those plants. The objective is to minimize total fixed and shipping costs. Constraints are that each customer's demand must be satisfied and that each plant cannot supply more than its capacity if it is open. Applications of the CFLP include location and distribution planning, lot sizing in production planning (Pochet and Wolsey, 1988), and telecommunication network design (Kochmann and McCallum, 1981; Mirzaian, 1985; Boffey, 1989; Chardaire, 1999).

Numerous heuristic and exact algorithms for the CFLP have been proposed in the literature. Classical heuristics apply ADD, DROP and interchange moves in conjunction with dominance criteria and approximations of the cost change caused by a move (Kuehn and Hamburger, 1963; Khumawala, 1974; Jacobsen, 1983; Domschke and Drexl, 1985; Mateus and Bornstein, 1991). Tabu search procedures have been developed for related problems

B. Fleischmann et al. (eds.), *Distribution Logistics*

like the p-median problem (Rolland et al., 1996) and the CFLP with single sourcing (Delmaire et al., 1999). Based on a rounding and filtering technique proposed by Lin and Vitter (1992), Shmoys et al. (1997) derive new approximation results for the metric CFLP. Korupolu et al. (1998) show that simple local search heuristics also give polynomial constant factor approximation algorithms for this problem. Magnanti and Wong (1981) and Wentges (1996) apply Benders' decomposition and show how to derive strong Benders' cuts for the CFLP. Polyhedral results for the CFLP have been obtained by Leung and Magnanti (1989), Aardal et al. (1995) and Aardal (1998). Aardal (1998) uses these results in a branch-and-cut algorithm for the CFLP. A variety of solution approaches for the CFLP, however, use Lagrangean relaxation. Lagrangean relaxation of the demand constraints with or without addition of an aggregate capacity constraint or another surrogate constraint is considered in Geoffrion and McBride (1978), Nauss (1978), Christofides and Beasley (1983), Shetty (1990), and Ryu and Guignard (1992). Beasley (1988) relaxes both the demand and capacity constraints; the resulting relaxation is, therefore, no stronger than the LP relaxation. Van Roy (1986) employs Lagrangean relaxation of the capacity constraints and cross decomposition in order to solve the resulting Lagrangean dual.

With the exception of Van Roy's (1986) cross decomposition algorithm, Lagrangean relaxation approaches for the CFLP generally use subgradient optimization in order to obtain an approximate solution to the Lagrangean dual. However, for solving larger and/or more difficult instances of the CFLP the knowledge of an exact solution of the corresponding master problem can be advantageous. Firstly, this gives an improved lower bound and, secondly, the knowledge of a fractional optimal solution of the primal master problem can be exploited to devise (better) branching decisions in the framework of a branch-and-price algorithm. In this paper, column generation is, therefore, employed in order to obtain exact solutions to the master problem. The approach is based on relaxing the demand constraints in a Lagrangean manner, and a hybrid mixture of subgradient optimization and a "weighted" decomposition method is applied for solving the master problem. Furthermore, the column generation procedure is embedded in a branch-and-price algorithm for computing optimal solutions to the CFLP.

This paper is organized as follows. The next section summarizes important Lagrangean bounds for the CFLP. Section 3 describes the column generation method, that is the master problem and the pricing subproblem, the employed method for stabilizing the decomposition as well as the branch-and-price procedure. Section 4 gives computational results which have been obtained for a set of problem instances with up to 200 potential plant locations and 500 customers. Finally, the findings are summarized in section 5.

2 Lagrangean Relaxation

Mathematically, the CFLP can be stated as the following linear mixed-integer program:

$$
\begin{aligned}
Z = \min \; & \sum_{k\in K}\sum_{j\in J} c_{kj}x_{kj} + \sum_{j\in J} f_j y_j & & \text{(O)}\\
\text{s.t.} \; & \sum_{j\in J} x_{kj} = 1\,, & \forall\, k \in K & \quad \text{(D)}\\
& \sum_{k\in K} d_k x_{kj} \le s_j y_j\,, & \forall\, j \in J & \quad \text{(C)}\\
& \sum_{j\in J} s_j y_j \ge d(K) & & \quad \text{(T)}\\
& x_{kj} - y_j \le 0, & \forall\, k \in K\,,\ \forall\, j \in J & \quad \text{(B)}\\
& 0 \le x_{kj} \le 1\,,\ 0 \le y_j \le 1\,, & \forall\, k \in K\,,\ \forall\, j \in J & \quad \text{(N)}\\
& y_j \in \{0,1\}\,, & \forall\, j \in J\,, & \quad \text{(I)}
\end{aligned}
$$

where K is the set of customers and J the set of potential plant locations; c_{kj} is the cost of supplying all of customer k's demand d_k from location j, f_j is the fixed cost of operating facility j and s_j its capacity if it is open; the binary variable y_j is equal to 1 if facility j is open and 0 otherwise; finally, x_{kj} denotes the fraction of customer k's demand met from facility j. The constraints (D) are the demand constraints and constraints (C) are the capacity constraints. The aggregate capacity constraint (T) and the implied bounds (B) are superfluous; they are, however, usually added in order to sharpen the bound if Lagrangean relaxation of constraints (C) and/or (D) is applied. Without loss of generality it is assumed that $c_{kj} \ge 0 \;\forall\; k, j$, $f_j \ge 0 \;\forall\, j$, $s_j > 0 \;\forall\, j$, $d_k \ge 0 \;\forall\, k$, and $\sum_{j\in J} s_j > d(K) = \sum_{k\in K} d_k$.

Lagrangean relaxation approaches for the CFLP relax at least one the constraint sets (D) or (C); otherwise the Lagrangean subproblem has the same complexity as the CFLP itself. Cornuejols et al. (1991) examine all possible ways of applying Lagrangean relaxations or Lagrangean decompositions to the problem consisting of constraints (D), (C), (T), (B), (N), and (I). Following their notation, let

- Z_R^S denote the resulting lower bound if constraint set S is ignored and constraints R are relaxed in a Lagrangean fashion, and let
- Z_{R_1/R_2} denote the bound which results if Lagrangean decomposition is applied in such a way that constraints R_2 do not appear in the first subproblem and constraints R_1 do not appear in the second subproblem.

Regarding Lagrangean relaxation, Cornuejols et al. (1991, Theorem 1) show that

$$Z^{BI} \le Z^{I} \le Z_C^{T} \le Z_C \le Z\,, \tag{1}$$
$$Z^{I} \le Z_D \le Z_C\,, \tag{2}$$
$$Z^{BI} \le Z_C^{B} \le Z_D\,. \tag{3}$$

Furthermore, they provide instances showing that all the inequalities above can be strict. The subproblem corresponding to Z_D can be converted to a knapsack problem and is solvable in pseudo-polynomial time. Therefore, bounds inferior to Z_D seem not to be interesting. Furthermore, as computational experiments show, Z_C^T is usually not stronger than Z_D. This leaves Z_D and Z_C as candidate bounds.

With respect to Lagrangean decomposition, Cornuejols et al. (1991, Theorem 2) proof that

$$Z_{C/D} = Z_{C/DB} = Z_{C/DT} = Z_{C/DBT} = Z_C\,, \tag{4}$$
$$\max\{Z_C^T, Z_D\} \le Z_{D/TC} \le Z_C\,, \tag{5}$$
$$Z_{D/BC} = Z_{D/TBC} = Z_{TD/BC} = Z_D\,. \tag{6}$$

Since Lagrangean decomposition requires to solve two subproblems in each iteration and to optimize a large number of multipliers, Lagrangean decomposition should give a bound which is at least as strong as Z_D. The only remaining interesting bound is, therefore, $Z_{D/TC}$. As shown by Chen and Guignard (1998), the bound $Z_{D/TC}$ is also obtainable by means of a technique called Lagrangean substitution, which substitutes the copy constraints $x = x'$ by $\sum_k d_k x_{kj} = \sum_k d_k x'_{kj}$. Compared to the Lagrangean decomposition, this reduces the number of dual variables from $|K| \cdot |J| + |J|$ to $2|J|$.

In summary, interesting Lagrangean bounds for the CFLP are Z_D, Z_C, and $Z_{D/TC}$. Compared to Z_C, the computation of the bound $Z_{D/TC}$ requires to optimize an increased number of dual variables. Furthermore, one of the subproblems corresponding to $Z_{D/TC}$ is an Uncapacitated Facility Location Problem (UFLP) while the subproblem corresponding to Z_C is an Aggregate Capacitated Plant Location Problem (APLP). Since the bound $Z_{D/TC}$ is no stronger than Z_C and since an APLP is often not much harder to solve than an UFLP, the bound $Z_{D/TC}$ seems not to be more attractive than Z_C. Compared to Z_D, the computation of the bound Z_C requires to repeatedly solve a strictly NP-hard subproblem, while the subproblem corresponding to Z_D is decomposable and solvable in pseudo-polynomial time. The column generation and branch-and-price procedure described in the following sections is, therefore, based on the Lagrangean relaxation of the demand constraints (D).

3 Column Generation Method

In the following subsections, the employed column generation scheme is described in detail. As shown in Section 3.1 the approach is based on relaxing the demand constraints (D) in a Lagrangean manner. Due to convergence problems of standard Dantzig-Wolfe decomposition, methods for improving convergence are required in order to solve the master problem efficiently. These methods are explained in Section 3.2. Finally, the branch-and-price algorithm for computing optimal solutions is outlined in Section 3.3.

3.1 Master Problem and Pricing Subproblem

Consider the mathematical formulation of the CFLP given by (O), (D), (C), (T), (B), (N) and (I). Dualizing constraints (D) with multipliers η_k, $k \in K$, gives the Lagrangean subproblem

$$Z_D(\eta) = \sum_{k \in K} \eta_k + \min_{x,y} \sum_{k \in K} \sum_{j \in J} (c_{kj} - \eta_k) x_{kj} + \sum_{j \in J} f_j y_j \quad \text{s.t.: (C), (N), (I), (B), (T)}\,. \tag{7}$$

It is easy to show, that optimal Lagrangean multipliers η^{opt} can be found in the interval $[\eta^{\min}, \eta^{\max}]$, where

$$\eta_k^{\min} = \min\{c_{kj} \,:\, j \in J \setminus \{j(k)\}\} \geq 0\,,\ c_{kj(k)} = \min_{j \in J} c_{kj}\,, \eta_k^{\max} = \max_{j \in J} c_{kj}\,.$$

Furthermore, it is well-known that (7) can be reduced to a knapsack problem (see, e. ,g., Sridharan (1993)). To this end, define

$$v_j = \max_x \Big\{ \sum_{k \in K} (\eta_k - c_{kj}) x_{kj} \,:\, \sum_{k \in K} d_k x_{kj} \leq s_j\,,\ 0 \leq x_{kj} \leq 1 \;\forall\, k \in K \Big\} \tag{8}$$

in order to obtain

$$Z_D(\eta) = \eta_0 + \sum_{k \in K} \eta_k\,,$$

where

$$\eta_0 = \min_y \Big\{ \sum_{j \in J} (f_j - v_j) y_j \,:\, \sum_{j \in J} s_j y_j \geq d(K)\,,\ y_j \in \{0,1\} \;\forall\, j \in J \Big\}\,. \tag{9}$$

In (8) and (9) constraints (B) are taken into account by setting $x_{ij} = 0$ if $y_j = 0$ holds in an optimal solution to (9).

The Lagrangean dual of (7) is to maximize the Lagrangean function $Z_D(\eta)$ over the set $[\eta^{\min}, \eta^{\max}]$. Let $\{y^t \,:\, t \in \mathcal{T}^y\}$ denote the finite set of all feasible solutions to the knapsack problem (9) and let $\{x_j^t \,:\, t \in \mathcal{T}_j^x\}$ denote the vertices of the set of feasible solutions to (8). For $t \in \mathcal{T}^y$ and $t \in \mathcal{T}_j^x$, define

$$F_t = \sum_{j \in J} f_j y_j^t \quad \text{and} \quad C_{tj} = \sum_{k \in K} c_{kj} x_{kj}^t\,.$$

Using these definitions and (8) as well as (9), the Lagrangean dual can be written as the linear program:

$$Z_D = \max \eta_0 + \sum_{k \in K} \eta_k \tag{10}$$

$$\text{s.t. } \eta_0 + \sum_{j \in J} y_j^t v_j \leq F_t \,, \qquad \forall\, t \in \mathcal{T}^y \tag{11}$$

$$\sum_{k \in K} x_{kj}^t \eta_k - v_j \leq C_{tj} \,, \quad \forall\, j \in J \,,\ \forall\, t \in \mathcal{T}_j^x \tag{12}$$

$$v_j \geq 0 \,, \qquad \forall\, j \in J \tag{13}$$

$$\eta_k^{\min} \leq \eta_k \leq \eta_k^{\max} \,, \qquad \forall\, k \in K \tag{14}$$

$$\eta_0 \in \mathbb{R} \,. \tag{15}$$

Taking the dual of the above "dual master program" one obtains the so-called "primal master program", which is given by:

$$Z_D = \min \sum_{t \in \mathcal{T}^y} F_t \alpha_t + \sum_{j \in J} \sum_{t \in \mathcal{T}_j^x} C_{tj} \beta_{tj} + \sum_{k \in K} \left(\eta_k^{\max} \overline{p}_k - \eta_k^{\min} \underline{p}_k \right) \tag{16}$$

$$\text{s.t. } \sum_{t \in \mathcal{T}^y} \alpha_t = 1 \,, \tag{17}$$

$$\sum_{t \in \mathcal{T}^y} y_j^t \alpha_t - \sum_{t \in \mathcal{T}_j^x} \beta_{tj} \geq 0 \,, \qquad \forall\, j \in J \tag{18}$$

$$\sum_{j \in J} \sum_{t \in \mathcal{T}_j^x} x_{kj}^t \beta_{tj} + \overline{p}_k - \underline{p}_k \geq 1 \,, \quad \forall\, k \in K \tag{19}$$

$$\alpha_t \geq 0 \,, \qquad \forall\, t \in \mathcal{T}^y \tag{20}$$

$$\beta_{tj} \geq 0 \,, \qquad \forall\, j \in J \,,\ \forall\, t \in \mathcal{T}_j^x \tag{21}$$

$$\overline{p}_k, \underline{p}_k \geq 0 \,, \qquad \forall\, k \in K \,, \tag{22}$$

where the variables α_t, β_{tj} as well as $\underline{p}_k$ and $\overline{p}_k$ are the dual variables of the constraints (11), (12) and (14), respectively.

If the constraints (20) and (21) are replaced by $\alpha_t \in \{0,1\}\ \forall t \in \mathcal{T}^y$ and $\beta_{tj} \in \{0,1\}\ \forall j \in J$ and $\forall t \in \mathcal{T}_j^x$, an equivalent formulation of the CFLP is obtained: The problem consists in selecting exactly one plant subset $S^t = \{j \in J : y_j^t = 1\}$, $t \in \mathcal{T}^y$, with sufficient capacity and in choosing feasible flows x_j^t, $t \in \mathcal{T}_j^x$, from plants to customers in such a way that total costs are minimized, each customer's demand is met (constraints (19)) and that there are no flows from closed plants to customers (constraints (18)). The primal master program (16)–(22) is the linear relaxation of this equivalent integer reformulation.

The primal master program (16)–(22) has to be solved by means of column generation. To this end consider the master problem *restricted* to known column subsets $\overline{\mathcal{T}}^y \subset \mathcal{T}^y$ and $\overline{\mathcal{T}}_j^x \subset \mathcal{T}_j^x$ for all $j \in J$. Furthermore let $\overline{\eta}_0$, $\overline{v}$, and $\overline{\eta}$ denote an optimal dual solution of the restricted master problem. New columns x_j^h and y^h price out, if

$$\overline{v}_j < \sum_{k\in K}(\overline{\eta}_k - c_{kj})x_{kj}^h \quad\Rightarrow\quad \overline{v}_j < v_j(\overline{\eta}) \stackrel{\text{def}}{=} \max\Big\{\sum_{k\in K}(\overline{\eta}_k - c_{kj})x_{kj}^t \;:\; t\in \mathcal{T}_j^x\Big\}$$

and

$$\overline{\eta}_0 > \sum_{j\in J}(f_j - \overline{v}_j)y_j^h \quad\Rightarrow\quad \overline{\eta}_0 > \min\Big\{\sum_{j\in J}(f_j - \overline{v}_j)y_j^t \;:\; t\in\mathcal{T}^y\Big\}.$$

Since $v_j(\overline{\eta}) \geq \overline{v}_j \;\forall\; j \in J$, using $v_j(\overline{\eta})$ instead of $\overline{v}$ in order to price out columns y^h is generally preferable; it leads to an earlier detection of required columns y^h.

For large instances of the CFLP even the restricted master problem is usually quite large, and the effort required for iteratively (re-)optimizing the restricted master problem can be tremendous. In addition it is well known that the conventional Danztig-Wolfe decomposition approach suffers from bad convergence behavior (see e. g. Lemaréchal (1989)). Methods for stabilizing the decomposition and reducing the computational effort required are, therefore, essential in order to solve the master problem (16)–(22) efficiently.

3.2 Stabilizing the Column Generation

When Lagrangean relaxation is applied to a general mixed-integer programming problem $\min\{cx \,:\, Ax = b,\; x\in X\}$, the Lagrangean dual is to maximize the piecewise linear and concave function

$$\nu(u) = ub + \min\big\{(c-uA)x \;:\; x\in X\big\} = ub + \min\big\{(c-uA)x^t \;:\; t\in\mathcal{T}\big\}, \quad (23)$$

where $\{x^t \;:\; t\in\mathcal{T}\}$ is the set of all vertices of the convex hull of X (for simplicity it is assumed that X is nonempty and bounded). For a given known subset $\overline{\mathcal{T}} \subset \mathcal{T}$ of columns, the function

$$\overline{\nu}(u) = ub + \min\big\{(c-uA)x^t \;:\; t\in\overline{\mathcal{T}}\big\}$$

is an outer approximation of $\nu(u)$. The restricted dual and primal master problem is then given by

$$\overline{\nu}(u^h) \stackrel{\text{def}}{=} \max_u \overline{\nu}(u) = \max_{u_0,u}\big\{u_0 + ub \;:\; u_0 + uAx^t \leq cx^t \;\forall\, t\in\overline{\mathcal{T}}\big\} \quad (24)$$

$$= \min_{\alpha\geq 0}\Big\{\sum_{t\in\overline{\mathcal{T}}}(cx^t)\alpha_t \;:\; \sum_{t\in\overline{\mathcal{T}}}(Ax^t)\alpha_t = b,\; \sum_{t\in\overline{\mathcal{T}}}\alpha_t = 1\Big\}, \quad (25)$$

where $u_0 \stackrel{\text{def}}{=} \min\{(c-uA)x^t \;:\; t\in\overline{\mathcal{T}}\}$ and α_t is the dual variable corresponding to the dual cut $u_0 + uAx^t \leq c^t$ for $t\in\overline{\mathcal{T}}$. At each iteration of the standard column generation algorithm (Kelley, 1960; Dantzig and Wolfe, 1960), the restricted master problem (24) is solved and an optimal solution x^h of the Lagrangean/pricing subproblem (23) for fixed $u = u^h$ is determined. The outer approximation $\overline{\nu}(u)$ is then refined by adding h to the set $\overline{\mathcal{T}}$ of columns. In order to improve the convergence behavior of this approach, a number of different methods have been proposed in the literature.

In order to avoid large oscillations of the dual variables u, Marsten et al. (1975) put a box centered at the current point, say u^{h-1}, around the dual variables u and solve

$$\overline{\nu}_\delta(\overline{u}^h) = \max_u \{\overline{\nu}(u) \,:\, u^{h-1} - \delta \leq u \leq u^{h-1} + \delta\} .$$

The next iterate u^h is then found by performing a line search into the direction $(\overline{u}^h - u^{h-1})$.

Du Merle et al. (1999) generalize the boxstep method of Marsten et al. They allow the next iterate to lie outside the current box, but penalize violations of the "box constraints". For these purposes they use the perturbed (restricted) dual master program

$$\begin{aligned} \max\ & u_0 + ub - w^+\pi^+ - w^-\pi^- \\ \text{s.t. } & u_0 + uAx^t \leq cx^t , \qquad \forall\, t \in \overline{\mathcal{T}} \\ & \delta^- - w^- \leq u \leq \delta^+ + w^+ , \\ & w^-, w^+ \geq 0 . \end{aligned} \tag{26}$$

Du Merle et al. propose different strategies to initialize the parameters π^+, π^-, δ^+, δ^- and to adapt them in case that an optimal solution u^h of (26) improves (not improves) the best dual solution found so far or in case that u^h is dual feasible.

As Neame et al. (1998) show, the method of du Merle et al. can be viewed as a penalty method which subtracts the penalty function

$$P_1(u) = \sum_i \max\{0,\, \pi_i^+(u_i - \delta_i^+),\, \pi_i^-(\delta_i^- - u_i)\}$$

from the outer approximation $\overline{\nu}(u)$ in order to determine the next point. The method of du Merle et al. is, therefore, closely related to bundle methods (Lemaréchal, 1989; Carraresi et al., 1995; Frangioni and Gallo, 1999) which use a quadratic penalty function

$$P_2(u) = \nu(u^{h-1}) + \frac{1}{2\tau}\left\|u - u^{h-1}\right\|^2 ,$$

where $\tau > 0$ is a "trust" parameter and u^{h-1} the current point. In this case, the master program or direction finding problem is a quadratic program.

Let $\nu(u^b) = \max\{\nu(u^t) \, : \, t \in \overline{\mathcal{T}}\}$ denote the best lower bound found so far. Optimal dual variables u are then located in the set

$$L = \{(u_0, u) \, : u_0 + ub - w_0 = \nu(u^b) \, , \; u_0 + u(Ax^t) + w_t = (cx^t) \, \forall \, t \in \overline{\mathcal{T}} \, , \; w_0 \geq 0 \, , \; w_t \geq 0 \, \forall \, t \in \overline{\mathcal{T}}\} \, .$$

Select any point $(u_0^h, u^h) \in L$ with $w_0^h = u_0^h + u^h b - \nu(u^b) > 0$. If (u_0^h, u^h) is dual feasible, that is $w_t^h = cx^t - u_0^h - u^h(Ax^t) \geq 0 \, \forall \, t \in \mathcal{T}$, then

$$\begin{aligned} \nu(u^h) &= \min\{cx^t + u^h(b - Ax^t) \, : \, t \in \mathcal{T}\} \\ &= cx^{t^*} + u^h(b - Ax^{t^*}) \, , \text{ for some } t^* \in \mathcal{T}, \\ &= w_{t*}^h + u_0^h + u^h b \geq u_0^h + u^h b = w_0^h + \nu(u^b) > \nu(u^b) \, , \end{aligned}$$

and the best lower bound increases at least by w_0^h. Otherwise, the localization set L is reduced by adding a column $t^* \in \mathcal{T} \setminus \overline{\mathcal{T}}$ which prices out at the current point u^h. Thus, a method which selects in every iteration such a point $(u_0^h, u^h) \in L$ converges in a finite number of steps to an ϵ-optimal dual solution u.

Interior point decomposition methods choose a point $(u_0^h, u^h) \in L$ obeying some centrality property. The analytic center cutting plane method (ACCPM) (Goffin et al., 1992, 1993) selects the point (u_0^h, u^h) which maximizes the (dual) potential function

$$\Psi(w) = \sum_{t \in \overline{\mathcal{T}}} \ln w_t + \ln w_0$$

over L. This requires to solve the non-linear system

$$\begin{aligned} &\mu_0 w_0 = \zeta \, , \; \mu_t w_t = \zeta \;\; \forall \, t \in \overline{\mathcal{T}} \, , \; \sum_{t \in \overline{\mathcal{T}}} \mu_t = \mu_0 \, , \; \sum_{t \in \overline{\mathcal{T}}} \mu_t(Ax^t) = \mu_0 b \, , \\ &w_0 = u_0 + ub - \nu(u^b) > 0 \, , \; w_t = cx^t - u_0 - u(Ax^t) > 0 \;\; \forall \, t \in \overline{\mathcal{T}} \, , \end{aligned} \tag{27}$$

where $\zeta = 1$. If $(w_0^h, w^h, u_0^h, u^h, \mu_0^h, \mu^h)$ is a solution to the system above, then (u_0^h, u^h) and $\alpha^h = \mu^h / \mu_0^h$ gives a feasible solution to the restricted dual master (24) and primal master (25), respectively.

Instead of computing the analytic center, Gondzio and Sarkissian (1996) as well as Martinson and Tind (1999) propose to use points on the central path between the analytic center and an optimal solution of the restricted master program (24). For these purposes a centrality parameter $\zeta > 0$ not necessarily equal to 1 is used and iteratively adjusted. Finally, Wentges (1997) simply proposes to select the point

$$(u_0^h, u^h) = \gamma(\overline{u}_0^h, \overline{u}^h) + (1 - \gamma)(u_0^b, u^b) \in L \qquad (0 < \gamma \leq 1) \, , \tag{28}$$

where $(u_0^b = \nu(u^b) - u^b b, u^b)$ is the best dual solution found so far and $(\overline{u}_0^h, \overline{u}^h)$ is an optimal solution of the restricted dual master program (24).

The parameter γ is first set to 1 and declined to a given threshold value in subsequent iterations. The convex combination (28) generally does not lie in the vicinity of a central path; nevertheless, the method is somehow related to interior point methods.

Van Roy (1983) introduces the cross decomposition method which combines Dantzig-Wolfe decomposition and Benders decomposition, and Van Roy (1986) uses this method to compute the bound Z_C for the CFLP. In order to avoid the need of solving a master problem in every iteration, the cross decomposition procedure obtains new dual iterates u^h from dual solutions to the primal subproblem, where integer variables are kept fixed, and generates new primal solutions x by solving the Lagrangean subproblem. When a convergence test indicates that convergence can no longer be expected, a master problem has to be solved in order to enforce convergence. Since subproblems are often easier to solve than master problems, a reduction in computational efforts is expected. If the primal subproblem does, however, not produce good dual information, the cross decomposition procedure may neither yield a reduced number of calls to the oracle nor a reduced number of master problems solved compared to a (stabilized) Dantzig-Wolfe decomposition method.

Last but not least, subgradient optimization and Dantzig-Wolfe decomposition can be combined in various ways in order to improve convergence. Guignard and Zhu (1994) use a two-phase method, which takes an optimal solution of the restricted dual master program (24) as next iterate only if subgradient steps fail to generate new columns for a given number of subsequent iterations. The restricted master is solved in every iteration in order to use the objective value $\max_u \overline{\nu}(u)$ as (improved) estimator of $\max_u \nu(u)$ in a commonly used step length formula.

Compared to Dantzig-Wolfe decomposition, bundle methods and interior point decomposition methods like ACCPM usually succeed in significantly decreasing the required number of calls to the oracle, that is the pricing subproblem. However, in case of large master programs and the addition of multiple columns in each iteration, the computational effort required for computing the next dual iterates can be substantial (in case of bundle methods a quadratic program needs to be solved; interior point methods (re-)solve the nonlinear system (27) by means of Newton methods). Such methods seem, therefore, to be suitable in case of a difficult subproblem and a relatively small master program. In case of the CFLP and the bound Z_D, however, even the restricted master program is usually quite large, while the pricing subproblem is relatively easy to solve. In order to solve the master program (16)–(22) it seems, therefore, more adequate to use a method which reduces the necessary number of calls to the master program even at the expense of an increased number of calls to the oracle. This can be achieved by means of mixing subgradient optimization and the weighted Dantzig-Wolfe decomposition approach:

- Good approximations of optimal multipliers η are quickly found by means of subgradient optimization. Furthermore, subsets of different columns y^t ($t \in \mathcal{T}^y$) and x_j^t ($t \in \mathcal{T}_j^x$) generated during subgradient optimization can be stored and used to initialize the restricted master program.
- The weighted decomposition scheme usually significantly improves convergence of Dantzig-Wolfe decomposition. Furthermore, the method is easy to use and "just" requires to reoptimize a linear program. In case that the dual prices u^h obtained from the convex combination (28) are dual feasible, the best lower bound $\nu(u^b)$ found so far increases by $\gamma\big(\overline{u}_0^h - \nu(u^b)\big)$. This gives the chance for further improvements in the current best lower bound if the parameter γ is increased in small steps until new columns price out at the current dual prices given by the optimal dual solution $(\overline{u}_0^h, \overline{u}^h)$ of the master program. Such a line search into the direction of $\overline{u}^h$ is feasible from a computational point of view only if the pricing subproblem is relatively easy to solve.
- Since the weighted decomposition as well as interior point methods may give feasible dual points, convergence can slow down at the end of the procedure if no additional columns are generated and the lower bound is already close to the optimum. Some intermediate additional subgradient steps may, however, help to overcome this situation. In our implementation, a limited number of intermediate subgradient steps are, therefore, performed in case that the next dual iterate obtained by the convex combination (28) is not dual feasible. During this intermediate subgradient phase, columns which price out at the dual prices given by the optimal dual solution to the master program are added.
- Furthermore, simple heuristics are used during and at the end of the column generation for determining (improved) feasible solutions for the CFLP. Columns from improved feasible solutions are added the master program. This guarantees that the objective function value of the restricted master program is no larger than the objective function value of a feasible solution for the CFLP. Finally, simple Lagrangean probing methods are employed in order to fix binary variables y_j if possible.

The column generation procedure employed for solving the master program (16)–(22) can be summarized as follows:

Column generation procedure

<u>Phase 1 (subgradient phase)</u>

Step 0: Set $h = 0$, $\overline{\mathcal{T}}^y = \overline{\mathcal{T}}_j^x = \emptyset \; \forall j \in J$, $LB = 0$, $UB = \infty$, $\overline{Z}_D = \infty$, $\sigma_h = 2$, and $\eta^h = \eta^{\min}$.

Step 1: Solve the pricing subproblem (7) for $\eta = \eta^h$. Let (y^h, x^h) denote the corresponding solution and let v^h, $\eta_0^h = Z_D(\eta^h) - \sum_{k \in K} \eta_k^h$ denote the values of v and η_0 corresponding to η^h. Set $O = \{j \in J \; : \; y_j^h = 1\}$. If $Z_D(\eta^h) > LB$, then set:

$$LB = Z_D(\eta^h),\ (y^b, x^b) = (y^h, x^h),\ \eta^b = \eta^h,\ v^b = v^h,\ \text{and } \eta_0^b = \eta_0^h .$$

Otherwise, half σ_h if the lower bound has not improved for a given number H^* of subsequent steps (e. g. $H^* = 10$). If $\big(\min\{\overline{Z}_D, UB\} - LB\big)/LB \le \varepsilon$, go to Step 11.

Step 2: Set $\overline{\mathcal{T}}^y := \overline{\mathcal{T}}^y \cup \{h\}$ and $\overline{\mathcal{T}}_j^x := \overline{\mathcal{T}}_j^x \cup \{h\}\ \forall\, j \in O$.

Step 3: Solve the transportation problem with plant set O. If this gives a solution improving UB, update UB and record the solution in (y^B, x^B).

Step 4: Set $\eta_k^{h+1} = \max\{\eta_k^{\min}, \min\{\eta_k^{\max}, \eta_k^h + \theta_h g^h\}\}\ \forall\, k \in K$, where

$$g_k^h = 1 - \sum_{j\in J} x_{kj}^h \quad \text{and} \quad \theta_h = \sigma_h\big(UB - Z_D(\eta^h)\big)/\|g^h\|^2 .$$

Set $h := h+1$. If h exceeds the iteration limit H (e. g. $H = 100$), go to Step 5, else go to Step 1.

Step 5: For each $j \in J$ compute

$$\rho_j = \min_y\Big\{\sum_{l\in J}(f_l - v_l^b)y_l \ :\ \sum_{l\in J} s_l y_l \ge d(K)\,,\ y_j = 1 - y_j^b,\ y_l \in \{0,1\}\ \forall\, l \in J\Big\}.$$

If $UB \le \sum_{k\in K}\eta_k^b + \rho_j$, then fix variable y_j to value y_j^b. If any (additional) binary variable could be fixed this way, recompute $\eta^{\min}$, $\eta^{\max}$ and perform some additional subgradient steps, that is set e. g. $H := H + 5$ and go to step 1. Otherwise, continue with Step 6.

<u>Phase 2 (column generation phase)</u>

Step 6: Initialize the primal master problem with columns $\{y^t : t \in \overline{\mathcal{T}}^y\}$ and $\{x_j^t : t \in \overline{\mathcal{T}}_j^x\}$ for which

$$(1-\epsilon)F_t \le \eta_0^b + \sum_{j\in J} v_j^b y_j^t \quad \text{and} \quad (1-\epsilon)C_{tj} \le \sum_{k\in K}\eta_k^b x_{kj}^t - v_j^b \tag{29}$$

holds, using e. g., $\epsilon = 0.01$. Add columns $\{y^B\}$ and $\{x_j^B : y_j^B = 1\}$ from the feasible solution to the master problem.

Step 7: Solve the primal master problem (16)–(22) and obtain an optimal dual solution $(\overline{\eta}_0, \overline{\eta}, \overline{v})$ with objective value $\overline{Z}_D$. Remove all columns from the master which have been nonbasic for a certain number of subsequent iterations. If $\big(\overline{Z}_D - LB\big)/LB \le \varepsilon$, go to Step 11. Otherwise, set $h := h + 1$, $\eta^h = \gamma\overline{\eta} + (1-\gamma)\eta^b$, where $0 < \gamma < 1$, and go to Step 8.

Step 8: Solve the subproblem as in Step 1. If columns $\{y^h\}$ or $\{x_j^h\}$ price out at the current dual prices $(\overline{\eta}_0, \overline{\eta}, \overline{v})$, add them to the master problem and go to Step 9. Otherwise, go to Step 10.

Step 9: Apply a limited number ΔH of additional subgradient steps, that is repeat Step 3, Step 4 and Step 1 ΔH times in this order. During this intermediate subgradient phase, add all columns which price out at the current dual prices $(\overline{\eta}_0, \overline{\eta}, \overline{v})$ to the master problem. Furthermore, apply Step 5 whenever an improved feasible solution (y^B, x^B) is found. Return to Step 7 after completion of this intermediate subgradient phase.

Step 10: As long as no column prices out at the current dual prices $(\overline{\eta}_0, \overline{\eta}, \overline{v})$, increase γ in small steps and repeat Step 1 and Step 3 with $\eta^h = \gamma\overline{\eta}+(1-\gamma)\eta^b$. During this "line search" also apply Step 5 whenever an improved feasible solution is found. Afterwards, go to Step 7.

Step 11: Let $(\overline{y}, \overline{x})$ denote the computed optimal solution to the master problem in terms of the original variables. Sort $\overline{y}$ in decreasing order and open plants j in this order as long as total capacity is insufficient or $\overline{y}_j$ exceeds a given threshold value, e. g. 0.75. If this way a solution improving UB is found, update UB and record the solution in (y^B, x^B).

In order to further explain some of the above steps, it is appropriate to comment on the following points:

- The knapsack problems (9) were solved by means of the COMBO algorithm of Martello et al. (1999).
- The step size strategy employed in phase 1 is proposed in Ryu and Guignard (1992).
- The tolerance ε in Step 1 and Step 7 was set equal to $1/(2^{15} - 1)$.
- The restricted master becomes too large, if all different columns generated during the subgradient phase are added. Since (η_0^b, η^b, v^b) approximates an optimal solution of the dual master, it is expected that columns not meeting the selection criterion (29) will be nonbasic.
- In order to limit the size of the master problem, inactive columns have to be removed (Step 7). This reduces the computation time required for each master problem and generally increases the number of master problems to be solved. In our implementation, columns are removed, if they are inactive for 5 subsequent iterations.
- The parameter γ was set to a value of 0.2 for smaller test problems and to a value of 0.05 for larger problem instances. In Step 10, the parameter γ is incremented in steps of 0.05.
- The number ΔH of intermediate subgradient steps in Step 9 was set to 10. Compared to an application of the procedure without the use of intermediate subgradient steps, this halfed the required computation time for some of the larger test problems, although the number of calls to the oracle increased.
- Whenever an improved feasible solution (y^B, x^B) is found during the column generation phase, columns x_j^B $(j \in J)$ which price out at the current dual prices $(\overline{\eta}_0, \overline{\eta}, \overline{v})$ are added to the master program.

3.3 Branch-and-Price Procedure

Let $(\overline{\alpha}, \overline{\beta})$ denote the optimal solution to the master problem (16)–(22) computed by means of the column generation procedure described above. The corresponding solution $(\overline{y}, \overline{x})$ in terms of the variables of the original problem formulation is then given by

$$\overline{y} = \sum_{t \in \mathcal{T}^y} \overline{\alpha}_t y^t \quad \text{and} \quad \overline{x}_{kj} = \sum_{t \in \mathcal{T}_j^x} \overline{\beta}_{tj} x_{kj}^t \,.$$

If $\overline{y}$ is integral (which means that $\overline{\alpha}$ is integral), an optimal solution to the CFLP is reached. Otherwise, the column generation procedure has to be combined with an implicit enumeration method in order to obtain optimal solutions.

If at least one $\overline{y}_j$ is fractional, possible branching strategies are to branch on single variables y_j with $\overline{y}_j \in (0,1)$ or to branch on more complex constraints involving variables y_j. The simplest branching strategy is to require $y_j = 0$ on the "left" branch and to fix y_j to 1 on the "right" branch. This branching rule is relatively easy to implement. If y_j is fixed to a value $\delta_j \in \{0,1\}$, all present columns y^t with $y_j^t = 1 - \delta_j$ have to be removed from the master problem. Furthermore, it is easy to enforce the branching constraint $y_j = \delta_j$ in the pricing subproblem (7).

Branching on single variables y_j may have, however, disadvantages. The master problem (16)–(22) is usually degenerated and possesses multiple optimal solutions. It is, therefore, likely that there will be only a small progress in the lower bound if a branch on a single variable is performed. This may be avoided by means of branching on subsets of variables y_j. If $S \subset J$ is such that $0 < \sum_{j \in S} \overline{y}_j < 1$, it is possible to branch by introducing the pair of branching constraints

$$\sum_{j \in S} y_j \geq 1 \quad \text{and} \quad y_j = 0 \, \forall \, j \in S \,.$$

In our implementation the set S is determined by sorting $\overline{y}$ in non-decreasing order and including in this order plants $j \in J$ into the set S until $\sum_{j \in S} \overline{y}_j \approx 0.5$. The branching constraint $y_j = 0 \; \forall j \in J$ is easy to handle; it just requires to exclude plants $j \in S$ from consideration. However, the branching constraint

$$\sum_{j \in S} y_j \geq 1 \quad \Leftrightarrow \quad \sum_{t \in \mathcal{T}^y} \Big(\sum_{j \in S} y_j^t \Big) \alpha_t \geq 1 \qquad (30)$$

has to be added explicitly to the master problem. Each branching constraint of the type (30) contributes then to an additional dual variable, and the reduced costs of columns as well as the objective function of the pricing subproblem has to be adjusted accordingly. The general structure of the pricing subproblem is, however, not changed.

The branching rule suggested above is unbalanced in the sense that the constraint $y_j = 0 \ \forall j \in S$ is usually more restrictive than the constraint $\sum_{j \in S} y_j \geq 1$. One can expect that a branch with y_j fixed to zero for all $j \in S$ can be pruned quickly. In some sense, the above branching rule aims at generating an inequality of the type (30) which cuts off the fractional solution $(\overline{y}, \overline{x})$ and cannot be violated by an optimal solution to the CFLP. A possible drawback of this approach is that after the addition of several such constraints, the progress in the lower bound gets smaller and smaller.

The algorithm keeps a list of generated and processed subproblems (nodes of the enumeration tree) that were not fathomed. If this list is empty, optimality of the best feasible solution found is proved. Otherwise, the pending node with the smallest corresponding lower bound is selected for branching purposes.

In order to generate a subproblem, the following information is extracted from the father node. First the solution to the current master problem gives all active columns, which are stored together with an optimal basis. Furthermore, it is necessary to keep track of the best dual solution for the master problem; this information is required for performing steps of the weighted decomposition. After a node of the enumeration tree is selected, the branching constraint is added by deletion of the infeasible columns respective generation and addition of the new row $\sum_{j \in S} y_j \geq 1$. The storage requirements of this enumeration strategy are large; however, a "best-first" search strategy does usually contribute to a faster increase in the global lower bound than, for example, a depth-first search strategy.

4 Computational Results

The proposed branch-and-price procedure was coded in Sun Pascal and run on a Sun Ultra (300 MHz) to solve several test problems, which were generated according to the proposal of Cornuejols et al. (1991). Test problems for the CFLP generated this way are usually harder to solve than other problems of the same size. The test problems are divided into three different sets of problems which differ according to their tightness (ratio $r = \sum_j s_j / d(K)$ of total capacity and total demand). We used capacity tightness indices r of 3, 5 and 10, respectively. In each problem set, there are 5 problem types of each of the following sizes: 100 × 100, 200 × 100, 200 × 200, 500 × 100, and 500 × 200 where the first number is the number of customers and the second is the number of potential plant locations. Five problem instances have been generated for each given size and tightness index r. The transportation problems and the linear master problems arising in the computations of the bounds were solved by means of the procedures *CPXnetopt()* and *CPXprimopt()* contained in CPLEX's (1997) callable library (version 5.0).

In a first set of computational experiments, the computational effort required by the suggested column generation procedure was compared to the

Table1. LP relaxation solved using CPLEX

Size	SLP%	LP%	UB%	T_{SLP}	T_{LP}	T_H
			$r = \sum_j s_j/d(K) = 3$			
100×100	0.33	0.11	0.33	0.6	2.3	0.8
200×100	0.34	0.27	0.16	3.0	5.9	1.2
200×200	0.12	0.08	0.69	5.8	13.9	2.6
500×100	0.44	0.41	0.70	27.7	63.4	13.5
500×200	0.18	0.17	0.32	28.0	38.7	6.7
max	0.70	0.61	2.01	31.1	128.5	33.3
mean	0.28	0.21	0.44	13.0	24.8	5.0
			$r = \sum_j s_j/d(K) = 5$			
100×100	0.65	0.43	0.34	0.8	3.9	0.8
200×100	0.56	0.46	0.56	7.0	23.0	4.2
200×200	0.18	0.17	0.38	8.1	12.4	1.6
500×100	0.55	0.48	0.64	99.3	162.8	13.9
500×200	0.42	0.40	0.85	108.3	163.0	10.9
max	1.05	0.70	1.40	130.0	203.8	19.6
mean	0.47	0.39	0.55	44.7	73.0	6.3
			$r = \sum_j s_j/d(K) = 10$			
100×100	1.02	0.62	0.14	1.5	5.5	0.5
200×100	0.53	0.41	0.55	30.7	65.0	4.5
200×200	0.54	0.40	0.76	12.0	34.6	2.2
500×100	0.25	0.22	0.15	330.0	486.6	18.1
500×200	0.47	0.43	0.85	607.4	935.1	27.7
max	1.45	1.14	2.16	657.8	1138.2	50.6
mean	0.56	0.42	0.49	196.3	305.4	10.6
			Total			
max	1.45	1.14	2.16	657.8	1138.2	50.6
mean	0.44	0.34	0.49	84.7	134.4	7.3

effort required to solve the LP relaxation of the original problem formulation by means of a common linear programming software (CPLEX). This LP relaxation was solved in the following way: In a first step, the weak LP relaxation given by (O), (D), (C) and (N) is solved. Afterwards, violated implied bounds (B) are added until the strong LP relaxation is solved. Finally, in a third step, the LP relaxation was further strengthened by means of adding lifted cover inequalities, odd hole inequalities, flow cover inequalities and other submodular inequalities proposed by Aardal et al. (1995). In addition, simple rounding heuristics were applied in order to compute feasible solutions for the CFLP. Table 1 shows the results obtained with this procedure (averages over the five instances of each problem type). In Table 1, SLP% and LP% denote the percentage gap between the optimum value Z of the CFLP and the strong LP-bound Z^I and the computed LP-bound, respectively; UB% is the percentage deviation of the solution computed by the rounding heuris-

Table2. Results column generation procedure

Size	LB%	UB%	It_{LR}	It_M	Col_A	Col_{Tot}	T_{LR}	T_H	T_M	T_{Tot}
				$r = \sum_j s_j/d(K) = 3$						
100×100	0.05	0.00	183	11	239	518	0.8	0.3	0.4	1.7
200×100	0.23	0.00	232	25	418	807	1.6	1.2	1.0	4.0
200×200	0.05	0.00	220	18	370	896	2.8	1.7	1.5	6.3
500×100	0.40	0.27	575	56	1115	2618	12.2	11.5	28.6	52.8
500×200	0.16	0.02	300	24	984	2139	8.9	17.0	9.9	36.7
max	0.65	0.70	636	62	1210	2749	13.0	23.6	33.6	58.9
mean	0.18	0.06	302	27	625	1396	5.3	6.3	8.3	20.3
				$r = \sum_j s_j/d(K) = 5$						
100×100	0.19	0.06	206	12	285	523	0.9	0.2	0.4	1.6
200×100	0.44	0.26	395	53	404	925	3.3	1.8	3.0	8.4
200×200	0.12	0.06	217	18	410	821	3.0	1.8	1.1	6.5
500×100	0.55	0.88	915	100	1062	3091	19.7	18.6	99.3	138.4
500×200	0.40	0.37	631	68	1087	2772	25.7	21.2	39.6	87.6
max	0.73	1.95	1109	129	1196	3314	30.5	28.9	146.7	180.2
mean	0.34	0.33	473	50	649	1627	10.5	8.7	28.7	48.5
				$r = \sum_j s_j/d(K) = 10$						
100×100	0.23	0.00	323	26	253	585	0.8	0.1	0.9	1.9
200×100	0.46	0.34	473	55	300	953	4.0	1.8	7.1	13.0
200×200	0.21	0.00	364	44	472	935	3.5	0.8	1.9	6.5
500×100	0.24	0.19	1003	105	948	3188	16.5	29.8	190.3	237.5
500×200	0.47	1.09	901	91	1123	3312	42.1	35.5	172.3	251.7
max	0.75	1.50	1450	171	1246	3961	56.1	50.3	271.2	321.3
mean	0.32	0.32	613	64	619	1795	13.4	13.6	74.5	102.1
				Total						
max	0.75	1.95	1450	171	1246	3961	56.1	50.3	271.2	321.3
mean	0.28	0.24	463	47	631	1606	9.7	9.6	37.2	57.0

tic from optimality; T_{SLP}, T_{LP}, and T_H are the times in seconds required to compute the strong LP-bound, the LP-bound, and heuristic solutions.

Table 2 gives information about the computational effort required to solve the master problem (16)–(22) by means of the proposed column generation procedure. In this table, It_{LR} and It_M denotes the number of subproblems and master problems solved, respectively; Col_A is the number of columns in the last master problem, and Col_{Tot} is the total number of columns generated; T_{LR}, T_H, and T_M denote the computation times in seconds required for solving the subproblems, the transportation problems and the master problems, respectively; T_{Tot} is the total computation time in seconds; LB% and UB% denote the percentage deviation of the computed bound Z_D and heuristic solution from optimality, respectively. Table 2 shows that the number It_{LR} of pricing subproblems solved is large compared to the number T_M of master problems solved: The procedure tries to avoid the relatively large master problem at the expense of an increased number of calls to the oracle. The computation times were, however, out of the scope, if instead of a weighted decomposition approach standard Dantzig-Wolfe decomposition was used (which means to set the parameter γ in (28) to a value of 1). We also experimented with the stabilization method proposed by du Merle et al. (1999), using the same subgradient procedure in order to obtain a good guess for optimal multipliers. The computation times were, however, significantly larger than those obtained by means of the proposed column generation scheme. Compared to the lower bound produced by the linear programming approach (column LP% in Table 1), the bound Z_D is on average better than this LP-bound. Furthermore, the proposed column generation method consumes less computation time than the computation of a bound based on the linear relaxation and additional cutting planes (compare columns $\mathsf{T}_{\mathsf{Tot}}$ in Table 2 and T_{LP} in Table 1). For larger problems with a capacity tightness of 5 and 10 the column generation procedure even consumed less computation time than the computation of the strong LP-bound Z^I by means of a simplex algorithm (compare columns $\mathsf{T}_{\mathsf{Tot}}$ in Table 2 and $\mathsf{T}_{\mathsf{SLP}}$ in Table 1). This indicates that this bounding procedure should be useful in the framework of a branch-and-price procedure for solving larger problem instances; it provides strong bounds in relatively short computation times and, in contrast to subgradient optimization, also a fractional primal solution on which branching decisions can be based.

In a second set of computational experiments the branch-and-price procedure described in Section 3.3 was used to compute optimal solutions. The method was also compared to two other exact solution procedures. The code of the branch-and-price procedure is, however, still under development. Furthermore, comparing different exact optimization procedures on large problem instances is very time-consuming. We are, therefore, only able to show preliminary results on some selected test problems.

The first method used for computing optimal solutions is the CAPLOC algorithm of Ryu and Guignard (1992). CAPLOC is a depth-first search branch-and-bound procedure based on Z_D and subgradient optimization. Before branching at the top node, however, CAPLOC tries to fix as many y variables as possible by means of extensive Lagrangean probing. The second alternative exact solution approach just consists in solving the LP relaxation of the original formulation in the way described above and in passing the problem together with the generated cuts and the computed feasible solution to CPLEX's MIP optimizer. Table 3 compares the results obtained with CAPLOC and CPLEX for some selected test problems. In this table, the numbers in the column headed Problem show the number of customers, the number of potential plant sites, the capacity tightness index r and the number of the problem instance. Nodes is the number of nodes checked and $\mathsf{T}_{\mathsf{Tot}}$ the total CPU time in seconds. Furthermore, Iter and #TPs denotes the number of subgradient steps performed as well as the number of transportation problems solved by CAPLOC. As can be seen from Table 3, the CAPLOC algorithm clearly outperforms CPLEX for the test problems shown. (Although

Table3. Results obtained by means of CAPLOC and CPLEX

	CAPLOC				CPLEX	
Problem	Nodes	Iter	#TPs	T_{Tot}	Nodes	T_{Tot}
100×100-3-1	103	274	16	1.1	9	5.2
100×100-3-2	106	397	35	1.8	182	24.8
100×100-3-3	113	500	18	1.4	248	25.3
100×100-3-4	131	767	23	2.0	103	8.9
100×100-3-5	125	793	38	2.4	103	11.8
max	131	793	38	2.4	248	25.3
mean	116	546	26	1.7	129	15.2
100×100-5-1	133	703	32	1.9	548	55.6
100×100-5-2	106	398	24	1.4	33	10.5
100×100-5-3	154	1264	77	3.6	325	45.1
100×100-5-4	293	1643	101	4.9	750	121.4
100×100-5-5	252	1684	94	4.2	178	25.7
max	293	1684	101	4.9	750	121.4
mean	188	1138	66	3.2	367	51.7
100×100-10-1	103	257	10	1.1	9	10.6
100×100-10-2	102	248	13	1.1	89	12.9
100×100-10-3	118	498	22	1.5	214	31.0
100×100-10-4	139	800	31	1.9	120	23.1
100×100-10-5	102	247	12	1.1	6	10.6
max	139	800	31	1.9	214	31.0
mean	113	410	18	1.4	88	17.6
200×200-3-1	260	1123	42	14.1	1146	382.7
200×200-3-2	206	558	41	9.8	1559	779.0
200×200-3-3	3590	17704	731	190.4	3019	2066.9
200×200-3-4	275	1138	34	13.8	1915	1185.4
200×200-3-5	1158	6528	358	78.8	524	429.9
max	3590	17704	731	190.4	3019	2066.9
mean	1098	5410	241	61.4	1633	968.8
200×200-5-1	426	2294	212	29.1	1555	774.7
200×200-5-2	871	5069	595	65.1	2221	1208.3
200×200-5-3	351	1662	100	20.1	578	313.1
200×200-5-5	200	318	39	8.0	40	38.9
max	871	5069	595	65.1	2221	1208.3
mean	462	2336	237	30.6	1099	583.7
200×200-10-1	495	3132	225	29.4	1459	1055.4
200×200-10-2	205	463	37	8.3	850	1355.9
200×200-10-3	228	841	75	12.6	731	421.6
200×200-10-4	523	3085	194	28.0	1413	798.1
200×200-10-5	204	460	43	9.7	832	1428.4
max	523	3132	225	29.4	1459	1428.4
mean	331	1596	115	17.6	1057	1011.9
	Total				Total	
max	3590	17704	731	190.4	3019	2066.9
mean	384	1906	117	19.3	729	441.5

for a few of the largest test problems not shown in Table 3, CPLEX were faster than CAPLOC).

Finally, Table 4 gives the results obtained for the selected test problems by means of the branch-and-price procedure of Section 3.3. In this table, Nodes is the number of nodes enumerated and Depth is the maximum depth of the enumeration tree; $\mathsf{It}_{\mathsf{LR}}$ and It_{M} denotes the number of subproblems and master problems solved, respectively; $\mathsf{Col}_{\mathsf{A}}$ is the number of active columns in the last master problem solved, and $\mathsf{Col}_{\mathsf{Tot}}$ is the total number of columns generated during the search; T_{LR}, T_{H}, and T_{M} denote the computation times required for solving the subproblems, the transportation problems and the master problems, respectively; $\mathsf{T}_{\mathsf{Tot}}$ is the total computation time in seconds. A comparison of Table 4 and the right hand side of Table 3 shows that the branch-and-price procedure is usually faster than CPLEX. The procedure enumerates much less nodes than CPLEX. The branch-and-price procedure consumes, however, more time per node enumerated than CPLEX. This indicates that reoptimizing the master problem takes more time than reoptimizing the LP relaxation, although the computation time required for solving the initial master problem by means of the described column generation method is for larger problem instances usually smaller than the time required to solve the LP relaxation of the original problem formulation. The current implementation of the branch-and-price method is, however, still not competitive to CAPLOC (compare Table 3 and Table 4). To some extend, this might be due to the employed branching rule, which seems not to be a good choice. While the branch $\sum_{j\in S} y_j = 0$ is usually pruned quickly, the increase in the lower bound along the branch $\sum_{j\in S} y_j \geq 1$ often happens to be much too small. More flexible and effective branching strategies are required in order to improve the performance of the algorithm. In addition, the computational effort required per node has to be reduced. An increase of the optimality tolerance between the lower bound and the value of the master problem at higher levels of the enumeration tree as well as the application of some extra subgradient steps directly after branching might be helpful in this respect.

5 Conclusions

In this paper, a column generation algorithm for the exact computation of a common Lagrangean bound of the CFLP was introduced. Furthermore, the column generation method was used within a branch-and-price procedure in order to compute optimal solutions. The column generation algorithm is based on a mixture of subgradient optimization and a variant of a "weighted" Dantzig-Wolfe decomposition approach. Obtaining optimal solutions to the master problem by means of this method usually requires less computation time than the determination of a bound based on the LP relaxation of the original problem formulation. The proposed branch-and-price method outperforms a branch-and-cut method (CPLEX) based on the LP relaxation of

Table4. Results obtained by means of the branch-and-price method

Problem	Nodes	Depth	It_{LR}	It_M	Col_A	Col_{Tot}	T_{LR}	T_H	T_M	T_{Tot}
100×100-3-1	3	1	240	16	366	808	1.0	0.3	0.8	2.2
100×100-3-2	11	5	613	81	302	1924	2.0	1.6	2.3	6.0
100×100-3-3	9	4	483	53	423	1643	1.2	0.5	1.8	3.7
100×100-3-4	19	7	1005	146	240	3831	3.0	2.0	4.0	9.4
100×100-3-5	7	2	469	44	384	1641	1.5	0.7	2.0	4.4
max	19	7	1005	146	423	3831	3.0	2.0	4.0	9.4
mean	10	4	562	68	343	1969	1.7	1.0	2.1	5.1
100×100-5-1	21	7	1369	149	521	4326	3.0	1.4	9.4	14.3
100×100-5-2	5	2	411	36	274	1247	1.0	0.4	1.8	3.4
100×100-5-3	33	10	2372	283	353	6603	5.9	3.0	16.8	26.8
100×100-5-4	177	27	11661	1610	347	33347	29.3	17.2	74.7	134.4
100×100-5-5	33	8	1903	248	389	5387	4.1	2.5	10.2	17.6
max	177	27	11661	1610	521	33347	29.3	17.2	74.7	134.4
mean	54	11	3543	465	377	10182	8.7	4.9	22.6	39.3
100×100-10-1	3	1	360	31	249	588	0.7	0.1	1.2	2.1
100×100-10-2	3	1	614	48	256	1214	1.1	0.1	3.2	4.6
100×100-10-3	3	1	474	39	438	1249	1.1	0.2	2.4	3.8
100×100-10-4	25	7	2088	227	352	4453	3.0	0.4	15.3	19.5
100×100-10-5	3	1	389	37	329	835	1.0	0.1	2.1	3.3
max	25	7	2088	227	438	4453	3.0	0.4	15.3	19.5
mean	7	2	785	76	325	1668	1.4	0.2	4.8	6.6
200×200-3-1	27	9	1542	273	484	5661	10.6	10.4	12.0	34.0
200×200-3-2	19	6	1231	189	429	5489	10.6	8.1	15.0	34.7
200×200-3-3	165	22	12287	2105	566	42748	111.6	103.4	142.0	373.4
200×200-3-4	9	4	732	105	524	2591	5.8	3.7	6.7	16.7
200×200-3-5	49	12	3009	547	519	10281	24.8	23.2	27.1	77.4
max	165	22	12287	2105	566	42748	111.6	103.4	142.0	373.4
mean	54	11	3760	644	504	13354	32.7	29.8	40.6	107.2
200×200-5-1	55	12	6154	994	479	17842	38.2	27.3	77.6	147.7
200×200-5-2	185	15	21235	3384	458	54688	114.2	105.3	279.3	529.5
200×200-5-3	153	21	14935	2264	670	41571	84.6	44.6	199.8	348.9
200×200-5-5	5	2	491	56	464	1660	3.6	1.9	4.9	10.9
max	185	21	21235	3384	670	54688	114.2	105.3	279.3	529.5
mean	100	13	10704	1675	518	28940	60.1	44.8	140.4	259.4
200×200-10-1	161	26	25160	3480	591	61284	108.8	35.8	508.4	700.9
200×200-10-2	13	3	2185	289	424	4659	9.3	1.3	31.1	42.8
200×200-10-3	27	9	3942	534	622	8800	19.9	5.6	74.5	102.6
200×200-10-4	93	17	14116	1937	411	33451	60.5	20.6	285.7	382.9
200×200-10-5	47	15	6479	929	607	15416	31.1	6.0	119.5	162.1
max	161	26	25160	3480	622	61284	108.8	35.8	508.4	700.9
mean	68	14	10376	1434	531	24722	45.9	13.9	203.8	278.3
					Total					
max	185	27	25160	3480	670	61284	114.2	105.3	508.4	700.9
mean	49	9	4955	727	433	13473	25.1	15.7	69.1	116.0

the original problem formulation. The current implementation of the method is, however, not competitive to a branch-and-bound method based on Lagrangean relaxation and subgradient optimization (CAPLOC). In order to improve the performance of the algorithm, the computational effort required for reoptimizing the master problem has to be further reduced. In addition, more effective branching strategies than the one used here are necessary. One possibility is to use the suggested branching rule only at top levels of the enumeration tree (or even only as long as the branch $\sum_{j \in S} y_j = 0$ can be pruned) and to branch on single variables afterwards. The selection of the branching variable can be based solely on the fractional solution to the current master problem or even on "degradations", which are computed from the master problem for a selected range of binary variables with a large fractional part of the solution value. Furthermore, the incorporation of polyhedral cuts should give rise to further improvements.

Acknowledgment

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.

Beasley, J. E. (1988): An algorithm for solving large capacitated warehouse location problems. *European Journal of Operational Research*, 33:314–325.

Boffey, T. B. (1989): Location problems arising in computer networks. *The Journal of the Operational Research Society*, 40:347–354.

Carraresi, P. / Frangioni, A. / Nonato, M. (1995): Applying bundle methods to the optimization of polyhedral functions: An applications-oriented development. *Ricerca Operativa*, 25:5–49.

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.

Chen, B. / Guignard, M. (1998): Polyhedral analysis and decompositions for capacitated plant location-type problems. *Discrete Applied Mathematics*, 82:79–91.

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.

CPLEX Division, ILOG Inc. (1997): Using the CPLEX callable library.

Dantzig, G. B. / Wolfe, P. (1960): Decomposition principle for linear programs. *Operations Research*, 8:101–111.

Delmaire, H. / Díaz, J. A. / Fernández, E. (1999): Reactive GRASP and tabu search based heuristics for the single source capacitated plant location problem. *INFOR, Canadian Journal of Operational Research and Information Processing*, 37:194–225.

Domschke, W. / Drexl, A. (1985): ADD-heuristics' starting procedures for capacitated plant location models. *European Journal of Operational Research*, 21:47–53.

du Merle, O. / Villeneuve, D. / Desrosiers, J. / Hansen, P. (1999): Stabilized column generation. *Discrete Mathematics*, 194:229–237.

Frangioni, A. / Gallo, G. (1999): A bundle type dual-ascent approach to linear multi-commodity min cost flow problems. *Informs Journal on Computing*, 11:370–393.

Geoffrion, A. M. / McBride, R. (1978): Lagrangean relaxation to capacitated facility location problems. *AIIE Transactions*, 10:40–47.

Goffin, J.-L. / Haurie, A. / Vial, J.-P. (1992): Decomposition and nondifferentiable optimization with the projective algorithm. *Management Science*, 38:284–302.

Goffin, J.-L. / Haurie, A. / Vial, J.-P. / Zhu, D. L. (1993): Using central prices in the decomposition of linear programs. *European Journal of Operational Research*, 64:593–409.

Gondzio, J. / Sarkissian, R. (1996): Column generation with a primal-dual method. Technical Report 1996:6, LogiLab, HEC, Section of Management Studies, University of Geneva. `http://ecolu-info.unige.ch/~logilab/reports`.

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.

Jacobsen, S. K. (1983): Heuristics for the capacitated plant location model. *European Journal of Operational Research*, 12:253–261.

Kelley, J. E. (1960): The cutting-plane method for solving convex programs. *Journal of the SIAM*, 8:703–712.

Khumawala, B. M. (1974): An efficient heuristic procedure for the capacitated warehouse location problem. *Naval Research Logistics Quarterly*, 21:609–623.

Kochmann, G. A. / McCallum, C. J. (1981): Facility location models for planning a transatlantic communications network. *European Journal of Operational Research*, 6:205–211.

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.

Kuehn, A. A. / Hamburger, M. J. (1963): A heuristic program for locating warehouses. *Management Science*, 9:643–666.

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.

Leung, J. M. Y. / Magnanti, T. L. (1989): Valid inequalities and facets of the capacitated plant location problem. *Mathematical Programming*, 44:271–291.

Lin, J.-H. / Vitter, J. S. (1992): Approximation algorithms for geometric median problems. *Information Processing Letters*, 44:245–249.

Magnanti, T. L. / Wong, R. T. (1981): Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29:464–484.

Marsten, R. E. / Hogan, W. W. / Blankenship, J. W. (1975): The Boxstep method for large-scale optimization. *Operations Research*, 23:389–405.

Martello, S. / Pisinger, D. / Toth, P. (1999): Dynamic programming and strong bounds for the 0-1 Knapsack problem. *Management Science*, 45:414–424.

Martinson, R. K. / Tind, J. (1999): An interior point method in Dantzig-Wolfe decomposition. *Computers & Operations Research*, 26:1195–1216.

Mateus, G. R. / Bornstein, C. T. (1991): Dominance criteria for the capacitated warehouse location problem. *The Journal of the Operational Research Society*, 42:145–149.

Mirzaian, A. (1985): Lagrangian relaxation for the star-star concentrator location problem: Approximation algorithm and bounds. *Networks*, 15:1–20.

Nauss, R. M. (1978): An improved algorithm for the capacitated facility location problem. *The Journal of the Operational Research Society*, 29:1195–1201.

Neame, P. / Boland, N. / Ralph, D. (1998): A unifying framework for column generation stabilization methods. Working paper, Department of Mathematics, University of Melbourne. `ftp://ftp.hpc.uh.edu/pub/ipco98/neame.ps`.

Pochet, Y. / Wolsey, L. A. (1988): Lot-size models with backlogging: Strong reformulations and cutting planes. *Mathematical Programming*, 40:317–335.

Rolland, E. / Schilling, D. A. / Current, J. R. (1996): An efficient tabu search procedure for the p-median problem. *European Journal of Operational Research*, 96:329–342.

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.

Shetty, B. (1990): Approximate solutions to large scale capacitated facility location problems. *Applied Mathematics and Computation*, 39:159–175.

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.

Sridharan, R. (1993): A Lagrangian heuristic for the capacitated plant location problem with single source contraints. *European Journal of Operational Research*, 66:305–312.

Van Roy, T. J. (1983): Cross decomposition for mixed integer programming. *Mathematical Programming*, 25:46–63.

Van Roy, T. J. (1986): A cross decomposition algorithm for capacitated facility location. *Operations Research*, 34:145–163.

Wentges, P. (1996): Accelerating Benders' decomposition for the capacitated facility location problem. *Mathematical Methods of Operations Research*, 44:267–290.

Wentges, P. (1997): Weighted Dantzig-Wolfe decomposition for linear mixed-integer programming. *International Transactions in Operational Research*, 4:151–162.

Modelling Warehouses in Logistical Networks

Angela Bauer

Universität Erlangen-Nürnberg, Lehrstuhl für Betriebswirtschaftslehre, insb. Logistik, Nürnberg, Germany

Abstract. Today's logistical networks are characterized by a high complexity and lots of interdependencies within and between supply chains. To handle this, strategic Decision Support Systems (DSS) help to make product flows transparent and to get an impression of the consequences of alternative net configurations before a decision is implemented. Nowadays DSS mainly focus on transportation flows and neglect detailed aspects of warehouse nodes when configurating logistical networks. But especially the interactions between transport and warehouse flows are of high importance for plausible results that resist implementation in practice. Therefore an integrated modelling of warehouse and transport processes is necessary. In this paper the Modular Node Model is introduced. Influences and consequences of alternative network decisions on warehouse processes and flows are modelled as well as feed backs to the transportation flows. This rather descriptive model serves as groundwork for optimisation algorithms. The focus lies on identifying relevant influence and decision criteria and developing a method to model relevant interactions. "Optimisation" is only considered in terms of the scenario technique. The model approach decomposes complex strategic decision problems in smaller manageable subproblems. This is done by building separate modules that are connected to each other by interfaces. The practicability of the model is verfied on the basis of an example from the consumer goods industry.

1 The Necessity of Modelling Warehouse Processes in Logistical Networks - Purpose and Procedure of the Paper

Logistical networks of today's marketplaces can be described as very complex and dynamic. This development has started in the 90s and is still going on. The reasons are especially due to the internationalisation of economic relations and the expanding cooperations and mergers of companies. Companies have to deal with increasingly enlarging supply chains. Decisions have to be made from a comprehensive point of view and generally there is only a small time slot to find optimal decisions in order to meet the customers' needs and to be superior to the competitors.

For this reason the modelling and calculating of whole networks is of high importance. Strategic Decision Support Systems (DSS) are necessary to handle mass data and to optimise structures and cost. The modelling of interdependencies of decisions, flows and processes in networks allows a fast assessment and choice of alternative solutions.

B. Fleischmann et al. (eds.), *Distribution Logistics*

There already exist a large number of suppliers in the market of strategic DSS for network planning. But the main attention of the available tools and models is paid to the analysis and optimisation of the transport flows and processes. However, the activities and processes in the warehouse nodes are mostly considered as restrictive parameters only, e.g. as capacity limits. The cost that incur in nodes is mostly calculated by cost rates per 100kg or per shipment and an additional fixed cost rate per day and node where applicable. This means that nowadays strategic DSS for network planning primarily use detailed data of the transporting processes, whereas data from warehousing etc. in nodes is considered by global cost rates.

Experiences from practical projects show that this general way of modelling warehouse nodes when assessing the consequences of alternative net configurations is not sufficient. To give an example, the optimisation of the European network configuration of a big consumer goods company implicated savings of up to 13% (see Section 4). But the underlying modelling approaches based only on transport data and fixed cost rates for the warehouse nodes. They did not consider the effects of various bundling strategies of the transport flows on the structure and quantities of shipments in warehouses. But both structure and quantities of shipments have far reaching influences on all activities in the different warehouse nodes of a network and consequently the cost incurring there might vary to a high extent. This holds especially when, in alternative scenarios, the number or function of warehouses is changed or when the allocation of customers and products to different warehouses is questioned. For example, the elimination of a warehouse node in a network can result in a relocation of functions and quantities to other nodes. This again leads to modified package categories, lot and shipment sizes. The results achieved with today's transport oriented tools may provide only "half the truth". In the case of the company here, the promised potentials of 13% could not be implemented in practice, as the complex interdependencies between transport and warehouse flows had been oversimplified and incurring cost had been neglected. In fact a potential of 10% could be realized.

The given remarks show that an integrated modelling of warehouse and transport processes in logistical networks is necessary.

Purpose and Procedure of the Paper

The purpose of this paper is to introduce a model – "The Modular Node Model" -, with the help of which a first step in modelling the interactions between transport and warehouse processes in logistical networks in a descriptive way could be taken. Influences and consequences of alternative decisions on warehouse processes and flows are pictured as well as feed backs to the transportation flows. This rather descriptive model serves as groundwork for optimisation algorithms. Therefore a large number of observations and examinations of warehouse and transport processes in companies had been made and implemented in software interfaces and data structures. "Optimisation" is only considered in terms of the scenario technique. That means that the model considers the essential dependencies and shows the various results for alternating scenarios. Through trial and error, a stepwise improvement is possible. True optimisation is neglected in this paper.

In order to build a complete model in a systematical way a methodical approach has to be found. For this it is assumed, that the configuration and optimisation of logistical networks always consist of a multitude of decisions to be made. Therefore the methodical approach is based on the theory of Decision Analysis [Samson (1988); Keen, Scott Morton (1978); Keeney, Raiffa (1976); Howard (1980); Raiffa (1968)].

The concept of Decision Analysis is built along the classical process of decision making [Samson (1988), S. 21ff; Turban (1988); S. 5ff; Simon (1960), S. 2-3; Domschke, Scholl (2000), S. 24ff; Bamberg, Coenenberg (1996)]. Its purpose is to structure and support decision processes and to fill the phases of the classical problem solving process with content. Following the steps of Decision Analysis or the typical phases of problem solving processes, respectively, the methodical approach of this paper and consequently the "Modular Node Model" itself is based on the following steps [Bauer (2002), S. 51f]:

First of all the problem "warehouses in networks" is identified and the structure of the decision model is built in Chapter 2. Within this first step, relevant key decisions are constituted and influence and decision criteria are designed.

Chapter 3.1 deals with the second step, the model building itself. A general model approach is founded and particular model components are specified.

In a third step in Chapter 3.2 the assessing and last but not least the decision making has to be done. Cost and time parameters are provided and the complex influences and interactions are formulated.

In a last step in Chapter 4, the model is verified by an application in practice.

2 Identifying and Structuring the Problem of Modeling Warehouses

General Characteristics of Warehouses

Warehouses primarily follow two intentions: on the one hand, a minimum of service level has to be assured, that means, that the production process has to be supplied with the necessary materials and goods as well as the market demand has to be met in time. On the other hand this necessary service level has to be provided by a minimum of cost.

The *standard process in a warehouse* depends to a high extent on the branch in view, the objects to be stored and the functions the warehouse should provide. This means a huge amount of possible process alternatives.

However literature largely describes consistent subprocesses that allow a complete description of the various functions in warehouses and of a general or generic warehouse process [Kopsidis (1997), S. 138ff; Eichner (1995), S. 26ff; Bowersox, Closs (1996), S. 397; Lambert, Stock, Ellram (1998), S. 277]. On a global level these process elements are “incoming goods”, “storage”, “picking” and “outgoing goods”. This global level can be further detailed (Figure 1).

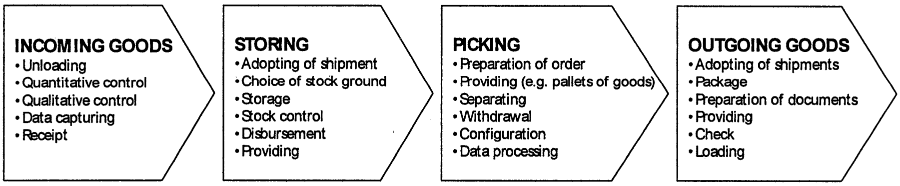


Fig. 1. Generic warehouse process in detail

In order to be able to model the dynamic interdependencies and interactions between alternative transport flows and network structures, the position and functions of warehouses in networks in general have to be discussed first.

Within a company’s network various types of warehouses with different functions appear. To describe the position of warehouses in logistical networks the order-to-payment process of a shipper serves as an example (Figure2).

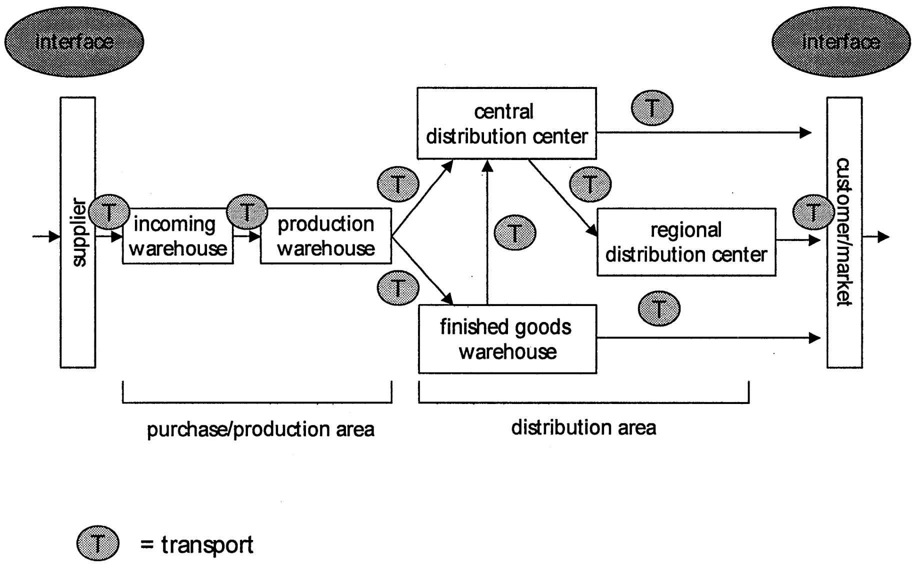


Fig. 2. Warehouses in the order-to-payment process of a shipper

The first contact with suppliers takes place in an incoming warehouse as interface. Here goods are received, temporarily stored if necessary and forwarded to further warehouses. This could be a production warehouse where goods are stored between the production processes or as finished goods not yet delivered. The next could be a warehouse for finished goods or a central warehouse for the distribution. Depending on the structure of the distribution network several regional warehouses might be passed through afterwards. It can be asserted, that a sequence of different types of warehouses can be found in companies or networks. This sequence differs depending on the required functions and network levels but basicly it consists of a typical chain of recurring types of warehouses which are connected over internal (fork lift or assembly line) or external transport elements. In each chain of warehouses, several characteristics of the branch concerned or special handling needs also have to be considerd. But each warehouse can differ in criteria like function, priority or cost parameters. To illustrate this by means of an example: regarding the distribution network of a company selling frozen products, every warehouse in the network has to fulfill all demands concerning the maximum temperature allowed and food in general. But whereas the incoming warehouse has to handle for example huge amounts of open pallets with fresh fish, which can only be stored for a very short time and require special handling, the central distribution warehouse might have to handle small customer commissioned packets of convenience food. Short durability is no longer the critical point but the delivery to customers on schedule.

Key Decisions

The key decisions that occur when configurating and optimising logistical networks must now be extracted. The initial situation is a suboptimal network in terms of structure or cost or the planned expansion of a company. This activates decision processes. The basic questions to be answered are the following [Bauer (2002), S. 68-73]:

1. How many levels should be established in a network, i.e. shall transports flow directly from source to receiver or are there several warehousing or transhipment levels to be passed through?
2. Which functions shall be provided on each level, i.e. to what extent are storaging, picking or turnover elements necessary?
3. How many nodes shall there be on each level and which throughput capacities shall be planned?
4. Where shall the node sites be located and how shall the allocation of customers and nodes to each other be designed?
5. How shall different product categories be allocated to nodes, for example are full or partial assortments preferred?

Along those five key decisions, alternative network scenarios can be built up as decision base.

Influence Criteria

To be able to assess and select alternative scenarios, relevant influence and decision criteria have to be extracted. The most important influence criteria when giving alternating answers to the five key decisions are the bundling effects, the structure and quantities of shipments and the dynamic and complexity of interdependencies and interactions within logistical networks.

Two kinds of *bundling effects* can be distinguished: on the one hand bundling means the transshipment and combining of fixed package categories for example from smaller vehicles having collected goods to big trucks for the long haule. On the other hand bundling is understood as separating package categories and picking and packaging new categories causing new shipment structures. In both cases the intention of bundling is the combining of smaller to bigger transport quantities in order to gain economic advantages. The last case, which is the changing of package categories, can highly effect the warehousing node itself. For example answering the key decisions 1,3 or 4 might lead to the fact that whole levels or warehouses are eliminiated. It can occur now, that one warehouse has to fulfil further bundling functions and therefore has to handle pallets instead of cartons. Investments may be necessary and problems, primarily in handling and executing might occur. The knowledge of different characteristics of bundling effects is of high importance. Corresponding rules have to be derived.

In addition to the bundling effects, the *structure and quantities of shipments* are an important influence criterion. If functions in nodes or even the levels of a whole network are changed when answering the five key decisions, the structure and quantities of shipments may shift. The structure of shipments is defined by the type of package, i.e. the amplitude, the weight, the volume, the package category, and the number of pieces per type of package.

The structure of shipments can at first not be defined as ordinal value. No definite proposition can be made whether a shipment has a "big" or a "small" structure. But a qualitative proposition is possible: The more a shipment is separated into smaller single packages, the more it has a small pieced structure. The less packages it contains, the more it is "big structured". The qualitative structure of shipments can be expressed as a quotient of "the number of handling units (m)" to "the number of single pieces (n)".

If the quotient equals 1, the structure of shipment is very small pieced, that means the shipment consists of as many packages as single pieces to be shipped. If the quotient equals 1/n, a big structure of shipment can be stated. All single pieces are bundled in one package unit, a pallet, for example. The influence criterion "structure and quantities of shipment" is clarified by an example: a company has two production plants, two product assortments and a two level distribution network structure. The distribution of finished goods is starting from the production plants A and B passing a central distribution center and from there coming about two regional distribution centers to the receiving customers. Neither the central distribution center nor the regional distribution centers provide a storing function. Goods are only bundled and turned over there. If the answers to the five key decisions now suggest a scenario in which the product sortiment from plant A should also be stored in the central distribution center in order to guarantee special service levels, the structure of shipments changes as follows: Plant B daily delivers packets as before, but plant A delivers full pallets to the central distribution center in larger time intervals. These are stored there and used, if required, for order picking. The delivery in small packets to the regional distribution centers and to the customers is bundled with the products of B. So the structure of shipments has changed in the production site A and in the zone of incoming goods in the central distribution center. This influences the execution process and cost. The other nodes and interfaces still handle the same structure of shipments.

Furthermore the *interdependencies* and interactions of alternative net configuration scenarios must be considered. If the number of levels in a network is changed or if functions of levels are varied, it must be made sure, that both functions and quantities are taken over by down- or upstream levels. If for example one warehouse node is eliminated, it has to be checked which nodes are capable to process the additional quantities. If levels or nodes are added to the network, the functions in existing nodes must be reduced if necessary. This dynamic has to be taken into account in order to assess the resulting reallocation of functions and quantities in a proper way.

Last but not least the *complexity of the network* as a whole is considered as influence criterion. If functions or quantities are added to particular nodes, those might reach capacity limits and investment is necessary. It could be that before reaching this capacity limit, a critical point of complexity is achieved where the existing facilities, the organization or the personnel are overextended. The error rate is increasing and quality and customer service become inadequate. Additionally a multi-level network in general causes higher efforts for coordination and administration.

Decision Criteria

To finish the problem structuring it is necessary to extract decision criteria which enable to assess and compare alternative network scenarios. The decision criteria are derived from the objectives of logistics itself [Klaus (1998), S. 322; Delfmann (1998), S. 361]: a demanded minimum of service level must be provided incurring a minimum of cost. Service level is defined as quality of distribution, which is measured by factors like running time, timeliness and dependability. So the decision criteria used to build the model to integrate dynamic transport and warehouse flows and processes are "cost" and "service level". To decide whether a scenario is better than another the answers to the five key questions have to be evaluated with respect to occurring cost and possible service levels. Normally there is a trade off between these two criteria. For example, closing one warehouse might lead to lower cost. But also the longer distances to the customer can mean a lower service level.

3 Fundamental Concept of Modelling Warehouses in Logistical Networks – "The Modular Node Model"

3.1 Model Concept and Selection of Components

Decomposition and Recombining of Suitable Modules as a General Approach

The basic idea of the model approach is to break up the complete network flow process in subprocesses within single nodes. The goal is to generate modules out of these subprocesses that can be combined flexibly. With these modules the illustration of any warehouse node should be possible, unattached by specific type of storage or by the multitude of possible functions and structures. The modules must fit together by interfaces and be further detailed in order to cover all necessary matters. By decomposing complex decision situations and afterwards recombinig subelements, the original complexity of the whole is reduced to manageable parts. These can be analysed and assessed separately. If the perspective would rest on the whole network, important aspects especially inside the nodes might be neglected. The decomposition allows the transparent illustration of complex interdependencies and interactions. The interfaces between the modules make sure however that consequences and results of alternative network structure decisions become visible.

Selection of Generic Subprocesses in Warehouses as Basic Modules

The selection of suitable subprocesses in warehouses as basic modules is founded on the generic warehousing process. This means that the process elements "incoming goods", "storage", "picking" and "outgoing goods" can be used as basic modules. Any warehousing node is connected to a transport element. To be more general the four basic models are renamed in "incoming goods", "balancing time gaps", "change of shipment order" and "outgoing goods". The "balancing time gaps"-module represents all kinds of buffering activites. The "change of shipment order" element stands for any turnover, picking, packing and bundling activities. Figure 3 shows the selected basic modules. Depending on the warehouse configuration a flexible combination of modules allows the necessary differentiation on a model level.

Necessity of Detailing the Basic Modules

The four presented basic modules are suitable as first approach. But to model the multitude of different kinds of storage and warehouses completely the four modules are still too abstract. A further detailing is necessary. In doing so attention must be paid to the right level of detailing. It must be deep enough to enable the illustration of different kinds of warehouses with respect to the critical differences.

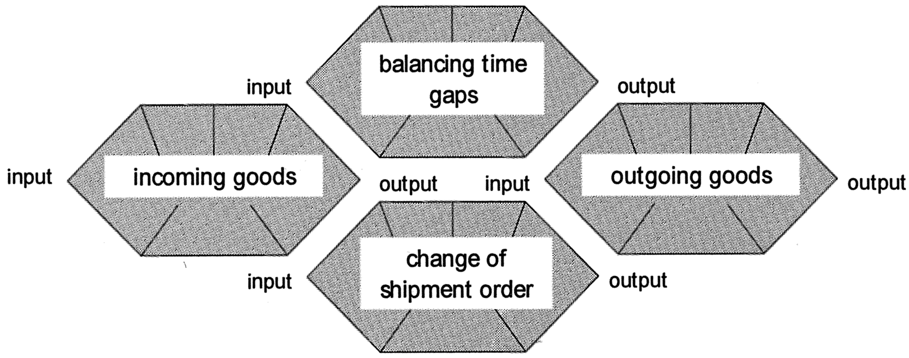


Fig. 3. The four basic modules

But it must also be general enough to be able to model different functions and specialities with only four modules. Each basic module has to be separated in sub activities which can be adapted to different types of warehouses at the critical points.

The module “incoming goods” is separated into the sub activities “unloading”, “identifying”, “quantitative control and check of delivery papers”, “qualitative control”, “data collection”, “receipt”, “editing and forwarding information” and “coding”. There is no varying of shipment structure or quantities. The incoming goods process serves only as a sluice for all shipments from outside when entering a certain node.

The module “balancing time gaps” is divided into the sub elements “adopting of shipments”, “choice of stock ground”, “taking into stock”, “data processing and receipt”, “stock control” and “disbursement” . This module is located inside a node, that means there is no direct interface to other nodes but only to other modules inside the same node.

The module “change of shipment order” consists of the subactivities “receiving of goods”, “order preparation”, “separating shipments”, “withdrawal of goods”, “identifying and picking”, “data processing and receipt”, “configuration and providing”.

The module of “outgoing goods” is assembled by the activities “receiving goods in the outgoing zone”, “packaging”, “preparation of documents”, “providing”, “outgoing check” and “loading”. Like the incoming goods process this module must occur in any warehouse node. It is also the interface to downstream nodes or transport elements.

Necessity of Completing the Basic Modules with Further Modules

A pure structural orientation of the modules along the static general warehousing process does not satisfy the purpose of a dynamic modelling of networks – a completion with corresponding modules is necessary.

First of all the registration of relevant attributes of the network frame along the five strategic key decisions is important. Furthermore the correlation between the network flow and the characteristics of the flowing objects must be described in a module. And last but not least the interfaces between the node model and the transport data are of high relevance.

The additional *module for the illustration of transports* has especially three special fields of application. First the transport module is needed, when for technical, organizational or capacity reasons inside a node additional transports become necessary that are not yet considered within the basic modules. For example this may be the case between the module of incoming goods and the module of balancing time gaps if, because of temporary bottlenecks, the incoming goods first have to be stored in the incoming zone before being transported to the proper storage zone. A second field of application is the transport between two separately modeled nodes that are located on the same site. The third field of application is external transports between nodes in different locations. The module of transports is further divided into the elements "loading goods", "transport" and "unloading".

The *module to charaterize the flowing objects* in a network serves as detailed information base. The focus lies on logistical attributes and their convertibility under the influence of dynamic interdependencies when varying the strategic network design. First of all, all products of a company have to be divided into groups with distinct attributes like "the same production plant", "specific package categories" or "temperature demands" etc. Subsequently the smallest handling unit of a product is defined with respect to its dimensions, its weight and possible special features like bulkiness or caution marks. Information about dimensions and weights of the next superior package categories are added, filled with a remark how many of the single units or smaller packages fit into the superior. This hierarchical order must be carried through from the smallest to the biggest package category.

The *module to illustrate the network structure frame* is the main item of the model. It constitutes the master network of a scenario along the five key decisions for net configuration and is therefore also called "scenario editor". The dynamic aspects of the model are fixed here. Beside the network structure typical shipment flows are pictured as well.

Figure 4 shows the schematic composition of a network with different levels and nodes in general. This network consists of three plants, three finished goods warehouses in level 2, two central distribution warehouses in level 3 and five regional distributions warehouses in level 4.

Any node is assigned with attributes like the executed functions or capacity limits. Finally the connections between the nodes themselves are defined. Hereby the rule applies that nodes are connected via the zones for incoming and outgoing goods. The connections themselves again can be filled with attributes like the allocation of customers and products to nodes, shipment structure and quantities, amount of required vehicles or the timing device of deliveries.

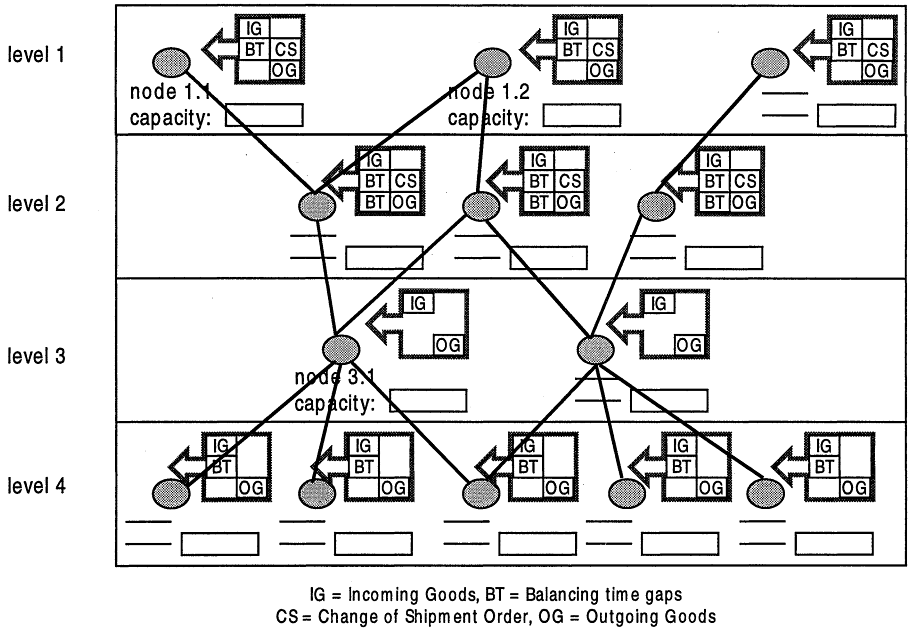


Fig. 4. Schematic composition of a network scenario

As soon as the network is defined with respect to its structure and connections the so called "master flows" are constituted. They describe the typical sequence of network levels and nodes passed through by the shipments and the corresponding conditions. For example shipments lighter than x kg pass the network via the levels 1,2,3,4 whereas shipments heavier than x kg go directly from level 3 to the receiving customer skipping level 4. As by weights master paths can also be defined by time slots, different countries or customers.

Last but not least the schematic master networks can be referred to geographical coordinates and so illustrated on a map.

If now levels or nodes are eliminated following the five key decisions this is registered by the scenario editor. The functions and quantities of the eliminated elements have to be reallocated.

3.2 Assessment Criteria

In this chapter the decision criteria "cost" and "service level" are precised as objective variables and the interdependencies and interactions are formulated.

"Resource Consuming" Modules versus" Information Providing" Modules

To calculate cost and service levels in a network any module with all subprocesses has to be examined separately. In doing so, it can be stated that there are two kinds of modules. The four basic modules “incoming goods”, “balancing time gaps”, “changing shipment order” and “outgoing goods” represent “real” logistical functions. The activities subsumed in these modules yield logistical value but they also consume cost and time resources. Input of personnel, technical equipment or capacities are required. That means that the activities of any of these four modules regularly “consume” resources, depending on the characteristics of the flowing objects and the shipment structures and quantities. Cost and service level are influenced by any single activity. The same holds for the transport module. The activities of this module also “consume” resources, whereas the modules “object character” and “master network – scenario editor” only provide information without consuming resources. This information is the groundwork for the other modules and they can have extensive influence on the development of cost and service levels in those modules. For this reason the following assessing functions are only developed for the four basic modules and the transport module.

Determination of Cost in “Resource Consuming” Modules

In any basic module some fixed and variable cost can be identified. These are for example fixed cost for buildings and properties, taxes, insurances, amortisation and energy. Those fixed cost for facilities are a given amount within a capacity limit and cannot be changed in the context of alternative scenarios. Fixed cost for technical equipment have to be considered as well. These are amortisations on investment, taxes, insurances and maintenance. In contrast to the facilitiy cost the technical fixed cost can change when a certain capacity limit of equipment is passed. The capacity limits must therefore be used as restrictions and in case be calculated with an additional fixed cost rate. The critical point here is how many shipments and quantities of the various types of package can be handled within one time slot. A parallel handling with several equipments of the same type must be taken into account. For example several fork lifts might be used to unload a truck and incur fixed cost for personnel. Thereby only personnel cost that incur completely independently from the passing shipment quantities is considered as fix. Cost for personnel in administration jobs and a minimum necessary crew are subsumed here. It is assumed, that in a warehouse facility at least two employees must be present to handle unexpected deliveries or orders, to do maintenance, care and cleaning jobs. This fixed cost for personnel is also considered not to be changed by alternative network scenarios.

Any activities that depend on shipment structures and quantities are charged with a variable cost rate per time unit. This cost rate results from the complete personnel cost of a special bracket and the corresponding working hours. It is assumed, that the more quantities are to be handled within one subprocess the more personnel time units are necessary. This variation of personnel cost along time units implies that personnel could be removed any time if shipment quantities decline. Although in practice this is not possible this way, the use of such a variable cost rate per time unit is reasonable because of the following:

- Normally a basic level of shipment quantities passing a warehouse is known approximately. So the basic personnel is adopted to the average time resources needed.

- If considerung strategic rearrangements or reconfigurations of networks with extensive consequences on shipment structure and quantities are in discussion a long-term adoption of the basic personnel is possible.
- If short-term fluctuations of quantities occur temporary employment or flexible working time models can be used.
- The basic personnel of a company is a self-regulating system. Constant labour surplus implicates either an adopting of capacities or the reallocation of personnel to other processes.

The determination of variable cost per subprocess is done by measuring the necessary time units needed per subprocess and multiplying it with the corresponding personnel cost rate per hour. Capacitiy restrictions need not to be calculated in this case.

Finally there are also variable cost that incur when using technical equipment. These are less dependant on time units but on the amount of pieces or handling units. Costs are generated not by time resources but by abrasion and energy use. With respect to the technical equipment, the need of time consumption per piece or handling unit is only relevant to determine capacity limits.

A last source of cost is complexity. If a certain limit of throughput quantities is surpassed, efficiency declines and a complexity cost rate must be charged in order to illustrate the escalating probability of mistakes and loss of time and cost.

Determination of Service Levels

To determine the service levels, that means especially the running times and the timeliness, the time units needed per subprocess in any module are relevant as well. Adding the time units needed across all modules, an overall running time of certain shipments can be calculated. By providing a fixed time frame for specific shipments, the keeping of time windows can be controlled. That means that collected data concerning time units for personnel and technical cost per sub process can also be used to calculate the relevant service levels.

Dependance of Ccost and Service Levels from Shipment Structure, Quantity and Complexity

The critical factor for determining cost and service levels as shown above is the consumption of time units. Time units again have strong correlations to the extracted influence criteria of the five key decisions "shipment structure and quantity". The more quantities are passed through warehouse node the more time units are "consumed". The "smaller" the structure of shipments is, i.e. the more single pieces are to be handled, the more time units are necessary, as in most subprocesses time units are determined by the amount of single pieces or handling units. For example the unloading of a certain quantity of shipments in form of the package type "pallet" from a truck can be managed much faster than the unloading of the same quantity in form of single packets.

Example of an Assessing Function for the Module "Incoming Goods"

The "Modular Node Model" consists of one assessing function for any subprocess in any module. This is assembled by fixed cost and by variable cost rates and needed time units. As example for such an assessing function the over all function for the module "incoming goods" is presented below.

Additionally to the fixed cost for facilities, technical equipment and personnel, all subactivities are provided with time units needed per package type. The total time over all types of package multiplied with the corresponding variable cost rates per time unit adds up to the variable personnel cost for the whole module of "incoming goods". Cost for technical equipment is calculated depending on the number of shipments. When calculating the cost for the subprocess "unloading" a speciality has to be taken into account. Here the number of handling units per type of package that can be transported within the scope of one way from the truck to the unloading place, by a fork lift for example, has to be included.

$$C = c_{fixF} + c_{fixT} + c_{fixP} + s \times (t_R + t_F) \times c$$
$$+ \sum_{i=A}^{Z} x_i \left[\left(\frac{t_{ULi}}{X_{i/process}} + t_{Ii} + p \times t_{QC} + t_{DC} + t_{COi} \right) \times c + \frac{v_{ULi}}{X_{i/process}} + v_{Ii} + p \times v_{QC} + v_{COi} \right]$$

c_{fix} =	Fixed cost: F=facilities, T=technique, P=personnel
x_i =	Number of handling units per type of package A-Z
$X_{i/process}$ =	Possible number of handling units per process per type of package
t_i =	Consumption of time per type of package in the zone of incoming goods (t_{ULi} per unloading process, t_{Ii} per identification, t_{QC} per quality control, t_{DC} per data capturing, t_{COi} per coding, t_R per receipt, t_F per forwarding)
v_i =	Technical cost rate per type of package per handling unit in the zone of incoming goods (v_{ULi} per unloading process, v_{Ii} per identification, v_{QC} per quality control, v_{COi} per coding)
c =	Personnel cost rate per time unit
p =	Percentage for snap check
s =	Number of shipments

3.3 The "Modular Node Model" – A Summary

As a conclusion "The Modular Node Model" can be illustrated as shown in Figure 5. Shipment data and the interface to the transport optimisation serve as input. Within the model, the scenario editor builds up master networks of different scenarios (structure component). The basis for this is the information from the object character module. The illustration of the warehouse nodes themselves is done by the four basic modules "incoming goods", "balancing time gaps", "change of shipment order" and "outgoing goods" and the additional transport module.

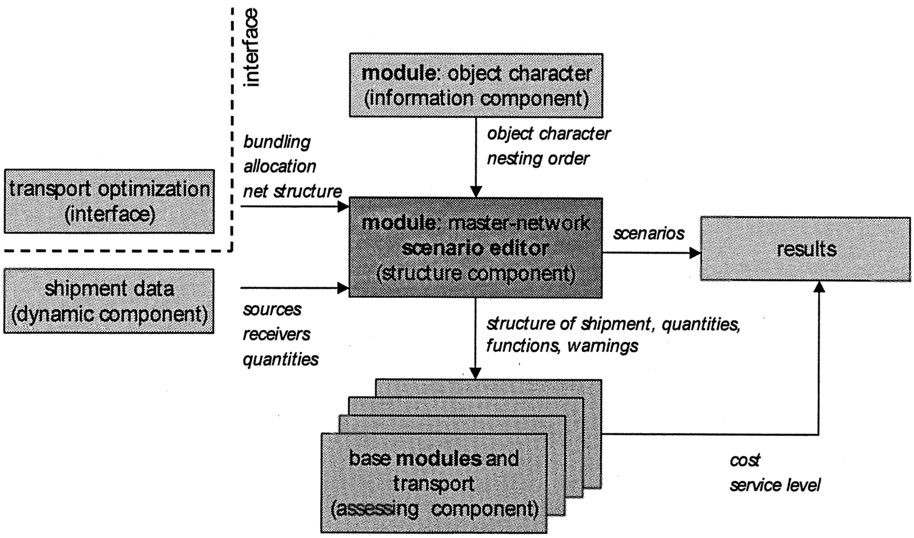


Fig. 5. "The Modular Node Model" - Survey [Bauer (2002), S. 167]

Using this model allows the picturing of any kind of warehousing node and the corresponding inderdependencies and interactions within a logistical network. This is verified in the following chapter.

4 An Example

To verify the practicability of "The Modular Node Model" a practical project of the consumer goods industry is introduced. The company in view sells high quality rather small shaped products and is running outlets all over the world. They asked for a detailed analysis of their transportation and distribution network and wanted to identify potentials of optimisation or further improvement of cost and service level. The first study had been run with focus on the transportation aspects exclusively. The company's European distribution network was run by three production plants and nine distribution centers. The distribution concept was country oriented, i.e. each country was served by a corresponding distribution center in this country, except for the Scandinavian countries, which were served by a single center. Any distribution center was served by all three plants and deliveries from overseas plants. So each center was fully assorted.

First of all, necessary input data about the present net structure was collected and prepared. In particular, this included shipment data with information about shipment date, weight, receiver, number of package and the distribution centers passed by. As the first study had considered only transportation aspects, the scenarios examined here were the following:

- Distribution is handled via one central distribution center, that serves all countries directly.
- Distribution is handled via three, four or five distribution centers that are all fully assorted.
- Distribution is handled in two stages via one central distribution center and several regional (national) distribution centers.
- Distribution is handled via two or three main distribution centers and several zero buffer transhipment points.
- In addition to the mentioned network scenarios the general strategy of bundling delivery days should be tested. That means for example a reduction of delivery days per customer from averaged three days to two or one day per week.

When examining these scenarios, warehouse cost were calculated as flatrate per 100 kg per warehouse. Differences occurred especially between countries. So the flatrate in Spain had been much cheaper than in Germany for example as wages were cheaper and as Germany additionally had a special bin-management.

The result of this first study had been, that, depending on the scenarios up to 13% savings were possible. The favourite scenario consisted of three main distribution centers and several zero buffer transhipment points. The bundling of delivery days increased the savings in any case but was not possible for all customers in terms of customer service guidelines.

During the elaboration of concrete implementation measures it became obvious that the changes that occur in the various locations due to a different distribution and warehouse structure differed extensively. Especially the different shipment structure and quantities that result from the reduction of the number of warehouses and the different distribution strategies had influence on cost and service level in the warehouses. It became clear, that in general, for all warehouses effective rules about consequences appearing in cost and service level are not sufficient. The calculated savings and potentials for each scenario could not in any case be verified. In some cases the different consequences in warehouses lead to much lower or also higher savings than anticipated on the basis of transportation aspects only. But also the ranking of different scnearios with regard to minimal cost can change. The reason for this is, that the location itself, the equipment, the capacities, the technique, the personnel and also certain strategies were already established and could not be adopted easily, if at all, to different shipment structures and quantities. For example the German distribution center had been oriented on the handling of small boxes each containing single orders. This holds for the incoming zone as well as for the order picking. If the shipment flow is changed over to full pallets, the handling cost increases tremendously as the whole warehouse is not adopted to pallet handling.

To be able to answer the question of how the distribution network could be optimised or improved a second study was established to examine the interdependencies between transportation strategies and flows and warehouse processes. The objective of this study had to a lesser extent been the optimisation of the network itself but to get the interdependencies and consequences of alternative scenarios pictured and transparent.

To fulfil the objective of this second project, the relevant decision and influence criteria had to be identified. After having observed several process activities in different warehouses and after having calculated time and cost parameters it became clear that the main decision criteria were cost and service level and that the most deciding influence criteria are the structure and quantities of shipments.

To apply the Modular Node Model a detailed list of package categories and object characteristics had to be completed. The various products were registered with their corresponding package alternatives and described with reference to attributes like weight, volume, product group, type of package and if necessary the number of containing units (e.g. a carton contains 20 single units of the product).

Also master networks with corresponding master paths and a detailed structure of warehouse nodes is built up. That means that any of the warehouse nodes of the active distribution network is described in detail by the modules of incoming goods, balancing time gaps, changing shipment order and outgoing goods. Furthermore any single activity in any warehouse is examined with respect to the time needed to fulfil this activity on one product and the corresponding cost. After that the warehouses are connected together on the basis of typical distribution strategies and shipment flows. If one alternative scenario is the elimination of several distribution centers, it can be illustrated by the scenario editor and calculated. The foundation for the calculation is the resulting shipment structure provided by the module to characterize the flowing objects and by the scenario editor as well as the cost and time data for each activity in each warehouse. From here a stepwise detailing of the scenario can start. The connection lines of the eliminated distribution centers as well as the functions and handled quantities in these nodes must be reallocated to the remaining nodes. After having established the scenario in the editor module the subprocesses of any node are investigated in detail and calculated with the assessment functions.

As a result it could be stated, that the potentials that had been proposed due to a simple optimisation of the transport structure with global warehouse parameters did not hold up when considering the dynamic interdependencies and interactions between transport and warehouse processes in the logistical network in detail. The determined potentials had only been “half the truth”. In reality rearrangements and investments had to be executed. The decision alternatives for the distribution network structure now appear in another light to the management. The average savings were reduced to 10% whereas the sequence of “best” scenarios also changed. The optimal number of warehouses stayed the same, but the location of these warehouses changed. Another result had been, that the Modular Node Model is able to picture the interdependencies and interactions between transport and warehouse flows in an appropriate way so that different scenarios could be calculated and that the decision maker can stepwise come to an improved solution.

5 Conclusions

The Modular Node Model helps to make the dynamic interdependencies and interactions between transport and warehouse processes transparent. The model approach decomposes complex strategic decision problems in smaller manageable sub problems. This is done by building separate modules that are connected to each other by interfaces. The modules are founded on the general warehousing process and complemented by additional modules for the illustration of flowing objects, for additional transports and for building master network structures. To apply the model a multitude of data and parameters is necessary. The practitioner has to weigh up the efforts for collecting the data in view of the promised savings potentials.

For the future, the model must be implemented as a software prototype to be able to test international and multi-modal networks. Moreover a data base should be established to find out typical cost rates and time units for special subprocesses. This would help to apply the model even if the available data base is insufficient. Above all, the results of this rather descriptive model should be developed into a quantitative optimisation model.

References

Bamberg, G. / Coenenberg, A. (1996): Betriebswirtschaftliche Entscheidungslehre. 9. Auflage, München

Bauer, A. (2002): Lagermodellierung für logistische Netze – Das „Modulare Knotenmodell" als Konzept zur dynamischen Abbildung von Lagern in Netzen. Hamburg

Bowersox, D. J. / Closs, D. J. (1996): Logistical Management. 3rd ed., New York, London

Delfmann, W. (1998): Netzwerkmodell. In: Klaus, P.; Krieger, W.: Gabler Logistik Lexikon, Wiesbaden, 361-362

Domschke, W. / Scholl, A. (2000): Grundlagen der Betriebswirtschaftslehre – Eine Einführung aus entscheidungsorientierter Sicht. Berlin, Heidelberg

Eichner, W. (1995): Lagerwirtschaft. Wiesbaden

Howard, R. A. (1980): An Assessment of Decision Analysis. *Operations Research* 28:4-27

Keen, P. G. W. / Scott Morton, M. S. (1978): Decision Support Systems: An Organizational Perspective. USA

Keeney, R. L. / Raiffa, H. (1976): Decisions with Multiple Objectives: Preferences and Value Tradeoffs. USA

Klaus, P. (1998): Logistikmanagement. In: Klaus, P.; Krieger, W.: Gabler Logistik Lexikon, Wiesbaden, 321-326

Kopsidis, R. M. (1997): Materialwirtschaft: Grundlagen, Methoden, Techniken, Politik. 3. Auflage, München, Wien

Lambert, D. M. / Stock, J. R. / Ellram, L. M. (1998): Fundamentals of Logistics Management, Illinois

Raiffa, H. (1968): Decision Analysis, Massachussetts

Samson; D. (1988): Managerial Decision Analysis, Illinois

Simon, H. A. (1960): The New Science of Management Decision, New York

Turban, E. (1988): Decision Support and Expert Systems, New York

Supply Chain Optimisation in Animal Husbandry

Jacqueline M. Bloemhof, Ciska M. Smeets and Jo A. E. E. van Nunen

Erasmus University Rotterdam, Rotterdam School of Management / Faculty of Business Administration, Rotterdam, The Netherlands

Abstract. The pig husbandry is an important economic sector. Major changes occurred in the last decade. As a result, farmers came together to introduce the 'Eco-label pig', meeting the strong consumer and governmental call for high quality, animal friendly and environmentally friendly food. The market for Eco-label food will grow enormously in the next years, asking for the development of an efficient and effective supply chain consisting of farmers, slaughter houses, wholesalers and retailers. We present some mathematical models to support decision-making and evaluation of a large number of growth scenarios, using cost minimization given a number of Eco-label conditions.

Keywords. Logistics, Network Configuration, Agricultural Logistics, Supply Chain Management.

1 Introduction

The pig husbandry in the Netherlands has changed considerably during recent years. Large scaling up took place leading to the disappearance of small firms and the increase of big firms. Further, the volume of the pig husbandry decreased as a whole with about 20 % in five years time and the image of the sector worsened due to diseases, restricting legislation and fodder scandals. To compete these threats, the pig husbandry introduced the Eco-label Pork, which is environmentally and animal friendly produced and of high quality.

The Dutch Foundation of Eco-labeling originated in 1992 as a co-production of government, producers, consumers, retail and environmental organizations. It is the Dutch competent body for the European Eco-label[1]. The aim is to stimulate animal and environmentally friendly production using a protected label. A product awarded with the Eco-label is much more friendly to the environment compared to most similar products. The Eco-label can be found on photocopying paper, labels, chairs, shoes, linoleum, writing pads, toilet paper, car polishing products, car-washes, cat litter, organic waste base plates, plants and flowers, bread, apples, pears, onions, barley, wheat, apple juice, potatoes and many other products.

In the mean time, the initiators of the Eco-label Pork have gained experience with the development of the Eco-label Pork as a socially sound, practically applicable and independent alternative for the usual pork production. At the moment, about 26 000 pigs a year are produced, treated and sold under Eco-label.

[1] http://europa.eu.int/comm/environment/ecolabel/

B. Fleischmann et al. (eds.), *Distribution Logistics*

Specific Eco-label requirements have to do with the farming and the logistics within the supply chain (e.g. restrictions on the maximum travel time from farmer to slaughter house). At the moment, the parties concerned with the Eco-label Pork aim to obtain a real sustainable supply chain as well as an increase in the volume of production and sales of the Eco-label Pork.

This asks for the development of an efficient and effective distribution network in the supply chain. The subject of this article is to obtain insight in the efficiency of distribution networks for the Eco-label Pork in the long run, by minimizing costs given specific Eco-label related restrictions.

Section 2 describes the problem framework, section 3 focuses on methods to analyze distribution systems. Section 4 introduces the model framework whereas section 5 contains the results of the scenario analysis. Finally, section 6 holds the conclusions.

2 Problem Description

The Eco-label organization De Hoeve b.v.[2] puts a lot of effort in the recruitment of pig-farmers for the Eco-label pigs supply chain. It expects a huge growth in the production of Eco-label pigs, which is necessary to obtain a profitable supply chain and a reasonable consumer price. The production of Eco-label pork uses capacity in existing facilities like trucks, slaughterhouses and wholesalers, although the products cannot be combined with pork without Eco-label. This situation may lead to a new assignment of facilities to products. Investments have to be made, capacity has to be reassigned, economies of scale get lost etc. These cost increases cannot fully be covered by price increases, as the customers' interest in green products will vanish very quickly. Therefore, reorganizing logistical processes in the supply chain to improve efficiency is very important.

The supply chain starts at the pig farmers. The pigs are distributed to slaughterhouses to process them into carcasses. The carcasses are distributed to wholesalers, and the Eco-label Pork is distributed to butchers (and in the future hypermarkets), where the pork is sold to customers (see Figure 1).

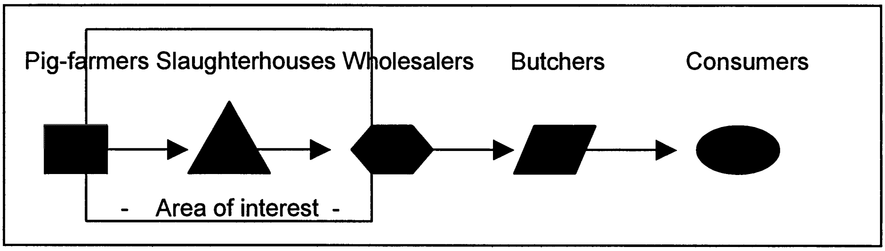


Fig. 1. Distribution structure for the Eco-label pigs

[2] http://www.dehoevebv.nl/

This research focused on Eco-label pigs, "ordinary" pigs from traditional farming were not taken into consideration. The production of the Eco-label pigs is at the moment quite small. The Eco-label organization wants the production to grow to one million pigs a year over the next five years. Changes in the distribution structure for Eco-label pigs will not cause large changes in the distribution structure for ordinary pigs, as the population of Eco-label pigs compared to the pig population in total will stay less than 6 percent.

We focus on the distribution between farmers, slaughterhouses and wholesalers. The production processes inside the pig-farmers, wholesalers and slaughterhouses are not included as they do not influence the distribution structure.

Processes at the butchers and the consumers are also not taken into account. Butchers are free to choose their own wholesalers. The Eco-label organization has marginal power to change these decisions. The Butchers union has obliged themselves to buy the Eco-label pork through a Covenant with the farmers, the slaughterhouses and the wholesalers. Therefore, whatever the production of the pork is, it will be sold. Finally, the consumers do not influence the choice of wholesalers and slaughterhouses and can be excluded from the problem area of interest.

Now, the distribution structure of the supply chain is rather straightforward because of the concentration of the production in the South of the Netherlands. Only one wholesaler and one slaughterhouse are part of the Eco-label pig supply chain[3]. Given the expected growth, the structure will become far more complicated.

Environmental effects of the supply chain occur mainly at the farms, e.g. ammonia emission in the air, manure excesses and land use (Blonk, 2001). Environmental effects of the distribution of pigs and carcasses involve (i) the length of the route and (ii) the type of vehicle used. We assume no difference in means of transport for traditional pigs and Eco-label pigs. The length of the routes will be part of the problem formulation.

The research objective is to:

Provide insight in the efficiency of a distribution network for Eco-label pigs in the long term, by minimizing relevant costs.

We focus on the following research questions:

- How is an efficient distribution network set up?
- Which methods and techniques fit to analyze distribution networks?
- Which conclusions can be drawn based on the efficiency analysis of the distribution network?

---

[3] http://www.milieukeur.nl/

3 Distribution Network

The distribution network for Eco-label Pigs can be designed from different perspectives, depending on the network structure that already exists and the objectives underlying the distribution network. Can Eco-label Pigs be distributed via an existing network or is a completely new network necessary. In this section we investigate some important location questions and motivate the choice for our solution procedure based on the characteristics of the problem.

3.1 Location Problems

We describe in this paragraph the important questions dealing with distribution networks. Designing a distribution network consists of a number of location decisions. These decisions contain the allocation of production, storage, capacity and transport related facilities. Location decisions can be split up in five important questions:

1. *Number:* How many locations should be placed?
2. *Location:* Where should the facilities be located?
3. *Facility role:* What role should each facility play? What processes are performed at each facility?
4. *Capacity allocation:* How much capacity should be allocated to each facility?
5. *Market and supply allocation:* What markets should each facility serve? Which supply sources should feed each facility?

All location decisions influence each other. That is why it is not possible to take one decision apart from the others (Chopra and Meindl, 2001).

3.2 Characteristics of the Eco-label Location Problem

The important questions for the Eco-label distribution network are:

- Which slaughterhouse(s) can best be used? How many, at which location?
- Which wholesaler(s) can best be used? How many, at which location?
- Which pig-farmers are allocated to which slaughterhouses? Which wholesalers are allocated to which slaughterhouses?

The number and location of pig-farmers are part of the data-input. The structure of the problem can be found in Figure 2.

Some important characteristics of the Eco-label Location problem are:

1. *Power:* The Eco-label organization has a coordinating role in the supply chain, but does not own (part of) the chain. Therefore, it cannot decide to open or close slaughterhouses or wholesalers as is possible for firms with locations owned.

2. *Sub network:* Location models typically involve locating new facilities with respect to existing facilities (Francis et al, 1983). In this case, traditional pork already makes use of the facilities and will do this in the future. Opening or closing a facility means using it for Eco-label pork or not. As a result, it is fairly easy to switch facilities.
3. *Product type:* The supply chain investigated consists for one part out of living material and for another part out of perishable products, so no stock costs arise in this case. Most location problems deal with tangible non-living products allowing for stocks.
4. *Amount:* Production facilities (pig-farmers) are many, but with a small production amount up to now. The location and size of future Eco-label pig-farmers is not known yet, so the individual coordinates of the pig-farmers cannot be used in the model. A solution could be clustering the pig-farmers, for example by region.

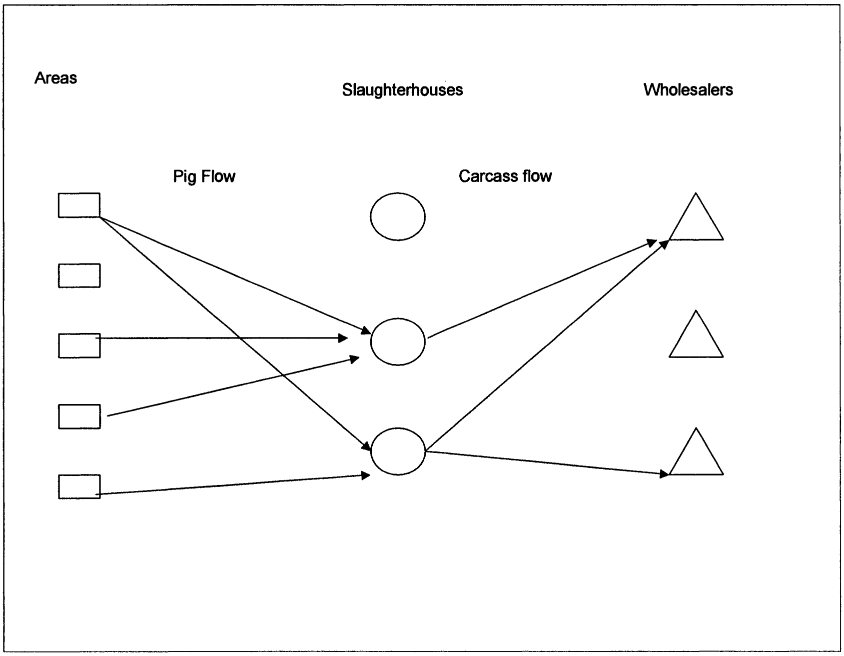


Fig. 2. Problem structure of the Eco-label distribution network.

5. *Push-pull:* In this location problem, the objective is not to fulfill demand (pull model), but to deliver production (push model). The butchers (i.e. the demand side of the supply chain) agreed in a covenant to take all Eco-label Pork produced in the supply chain. Therefore the supply side determines the flows in the network.

Next paragraph will describe the wide variety of location problems and their solution methods. Given this overview, we can decide on which type of location problem fits best for the Eco-Label Problem, and which solution method might be appropriate.

3.3 Overview of Location Problems

Many publications exist on location problems, due to the large diversity of the problem settings. This diversity proceeds from a variety of applications, underlying objectives and characteristics of the problem. The application can vary from location selection of factories, warehouses and distribution centers to allocating space in a factory to machinery, the cafeteria, tool room and toilets (Brandeau and Chiu, 1989).

The first distinction is related to the nature of the objective, which can be either qualitative or quantitative. Examples of qualitative factors are infrastructure, presence of labor, presence of land, nearness of markets and suppliers etc. (Chopra and Meindl, 2001) and (Sule, 2001). A well-known example of a qualitative method is the ranking method. With this method the managing team selects factors that they think are important for the location decision. Also, a maximum weight for each factor is determined on a scale from 0 to 100. Each location gets a score for each factor from 0 to the maximum weight. The location with the highest overall score is favorable. The qualitative methods are fairly easy to understand and also have as advantage the possibility of analyzing multiple factors at the same time. The disadvantage of these methods is the subjectivity through the influence of personal opinions of the management team (Heizer and Render, 2001). A quantitative method minimizes a mathematical formula consisting of quantifiable costs.

A qualitative method is suitable for developing distribution networks if the network is completely new or the network has to fulfill multiple objectives, for example with environmental burdening. This is essential with building facilities, if the environment, licenses and many strategic plans have to be taken into account. This method would be very worthwhile if no network structure exists, which is not the case for the Eco-Label location problem. Therefore, we focus on quantitative methods in the remaining of the overview.

Within quantitative methods, we distinguish continuous (or planar) and discrete location problems. A problem is continuous, if all locations in a certain space are possible, i.e. the possible locations form an infinite set (Plastria, 1995). A discrete location problem has a finite set, a given number of possibilities, for the possible locations (Barros and Labbé, 1994). Continuous methods can be very useful in developing a distribution network without any history. They can also be used if e.g. butchers and customer areas have to be located. However, the wholesalers and slaughterhouses in our problem setting already exist. There is no free choice of locations to build these facilities, so this specific problem has a final set of possible locations and is thereby a discrete problem.

Another characteristic of quantitative methods is the variety in mathematical objectives. Five categories can be mentioned (Sule, 2001):

1. p-Median problem
2. p-Center problem

3. Uncapacitated facility location problem (UFLP)
4. Capacitated facility location problem (CFLP)
5. Quadratic assignment problem

The p-median problem has the objective to place p facilities at p locations at minimum costs. The costs can be defined as value of time, money, distance, etc. This problem is also called minisum or Weber problem. The p-center problem or minimax problem has the objective to place p facilities at p locations minimizing the maximum distance from a facility to a point of demand. This type of objective is typical for locating emergency services, such as fire stations, police and hospitals (O'Kelly, 1995).

The uncapacitated facility location problem (UFLP) (Krarup and Pruzan, 1983) has the same objective as the minisum problem, but the cost formula contains a fixed cost component, i.e. also costs are related to the facilities themselves. Because the capacity of each facility is infinite, it is never optimal to allocate a specific demand to multiple facilities. The capacitated facility location problem (CFLP) (Cornuejols et al., 1991) is the same as UFLP, except for the capacity constraint.

The Quadratic assignment problem defines problems with n facilities, such as n machines with their own production flow, at n locations that have to be placed at the same time minimizing costs.

Different problem categories are suitable, depending on the cost structure. If all cost components deal with transportation, a p-Median or p-Center problem will satisfy. If some supply chain costs are related to facilities, the facility location problem suits best. If facilities have restricted capacities, due to the fact that these facilities already exist, the problem can be best defined as a capacitated facility location problem (CFLP).

A variety of different exact and heuristic solution approaches have been developed to solve location problems (Brandeau and Chiu, 1989). Many location problems have a discrete solution set and can be solved using integer and linear programming. If the problem has too many potentially optimal solutions or the location problem has an infinite number of potentially optimal solutions, heuristic solution techniques have to be used. The Eco-label location problem has about 20 possible locations for the slaughterhouses and about 20 possible locations for the wholesalers, making mixed integer linear programming a suitable tool to find the optimal solution.

The next characteristic is the number of levels of the location problem. A single-level location problem has to locate one level of locations, whereas other locations are fixed. If multiple facilities of different nature have to be located simultaneously, such as warehouses and factories (two-level), the problem is a multi-level problem (Bloemhof et al, 1996, Tcha and Lee, 1984).

A last characteristic is the number of periods in the problem. Location decisions are in general long term of nature, because of the investment costs. If an organization expects few changes in the market situation, the problem is often analyzed as a single period problem, i.e. with one set of input data. The counterpart of a single-period problem is a multi-period facility location problem (MFLP). This problem suits organizations with very variable market or cost conditions over time. They can make decisions on the number of facilities to be located, with what production, in each period for a known demand. These decisions are made up front, with a rolling horizon in which the values of the variables of the problem can change. An overview on exact and heuristic solution methods can be found in Canel and Sidhartha (1999). These authors also describe a combination of mixed-integer programming and dynamic programming to solve the MFLP. With this method, first all individual periods are optimized, then the overall solution is analyzed with dynamic programming for the whole time horizon.

3.4 Typology of the Eco-label Location Problem

The characteristics of the Eco-Label location problem lead to the following typology:

1. *Qualitative / Quantitative:* A selection is to be made out of existing wholesalers and slaughterhouses. Main objective is cost minimization, given a number of environmental conditions. Because of this, a quantitative method is suitable.
2. *Continuous / Discrete:* The problem has a final set of possible locations and is therefore a discrete problem.
3. *Mathematical objective:* The slaughtering costs belong to the supply chain costs. These costs are related to facilities, (0-1), and have a fixed part and a variable part. Both slaughterhouses and wholesalers have restricted capacities because these facilities already exist. Therefore, the problem is defined as a Capacitated facility location problem (CFLP), which can be formulated as a mixed integer linear programming (MILP) model.
4. *Single-level / Multi-level:* A selection of slaughterhouses and wholesalers has to be made. Pigs are transported from farm areas to slaughterhouses and then from slaughterhouses to wholesalers. As a result, the problem is a two-level problem.
5. *Single-period / Multi-period:* The research has a strategic character. The objective is an analysis of the distribution network in a time horizon of five years, so the problem is a multi-period location problem with five periods.

Summarizing, the Eco-label Location problem is a discrete two-level multi-period capacitated facility location problem.

4 Model Framework

The research objective is to develop an efficient distribution structure of the Eco-label Pigs supply chain, minimizing logistical costs. The model contains eco-label requirements to protect the environment and animal well being as follows. Firstly, only slaughterhouses and wholesalers in the Netherlands are examined. In traditional pig husbandry, a large amount of the living pigs are exported to Spain, Italy and other European countries to be slaughtered over there, and imported as pork to be consumed in the Netherlands. This occurs because farmers get a higher price abroad and it is cheaper to transport living pigs than pork that has to be cooled. Secondly, living pigs are allowed to be transported only 200 kilometres and are processed in the slaughterhouses early in the morning to avoid stress. In traditional pig husbandry, on average 500 kilometres is normal. This condition is modelled by adding penalty costs to the transportation costs.

The following elements of the modeling process must be considered to answer the general research questions of section 2, worked out in section 3.

- Making assumptions
- Preparing data for input in the model
- Modeling specific transportation and slaughtering costs

Notation

- Formulating the mathematical model

4.1 Assumptions

First, data has to be gathered on the amount of pigs in the Netherlands, the share of Eco-label pigs, and the spread over the farm areas. Secondly, assumptions have to be made concerning relevant costs. In the pig husbandry, living material and perishable products are transported from A to B. Stocking has to be avoided because an extended throughput time decreases the quality of the animal or the meat. Therefore, the supply chain processes can be qualified as cross-dock processes. Inventory is only for a couple of hours until a truck is available for transport and never more than 12 hours. Relevant costs are transportation costs and facility costs.

4.1.1 Preparing data

To evaluate long-term scenarios, a number of data has to be prepared in advance.

- *Regions.* For calculating the transportation costs from the farmers to the slaughterhouses and from the slaughterhouses to the wholesalers, a regional division of the Netherlands is necessary. Firstly, because we analyse future scenarios in which the precise locations of future farmers is not known. Secondly, the producing farmers themselves are presumably too small in size, especially in the first years of the forecast when the volume of the Eco-label pig population is not that high, and too many to deal with. Therefore it is reasonable to aggregate the producing farmers to about 40 areas of contracted municipalities.
- *Year productions.* In order to forecast growth scenarios for Eco-label Pigs, it is necessary to gather data on the current spread of pig production in the Netherlands. The year production of a region is calculated by multiplying the cage capacity in a region with a cycle factor (i.e. the number of times the cage is occupied on average during a year).
- *Forecast of population.* We assume an exponential growth in the production of Eco-label pork, based on information from the Eco-label organization. Starting, the production of Eco-label pigs is low, with a rather high entry-barrier. As the

production and sales of Eco-label pigs becomes more accepted, more and more farmers will switch to the Eco-label, decreasing the entry barrier. This so-called snowball effect can best be modelled by an exponential growth model, i.e.

$$Y^i_t = Y^i_0 \times g^t \tag{1}$$

with

Y^i_t the number of pigs in area *i* in year *t*

Y^i_0 the number of pigs in area *i* in year 0

g the growth rate (constant)

The number of Eco-label pigs in period 0 is assumed to be 26 000. If e.g. the goal is to arise at one million pigs in year 5, the growth rate is 2.07. At the moment only three areas in the Southeast of the Netherlands produce Eco-label pigs. Therefore, the regional spread of the traditional pig husbandry is used to assign values to the various regions in the first year[4].

4.1.2 Transportation costs

Transportation costs for pigs are calculated per load. They depend on the distance, the duration of load, unload and transportation, and the number of pigs to be transported. A load has on average about 100 pigs. Costs per load can be quantified as: 1.363 + 0.0056 * #kilometers of the outward journey. Further costs are penalty costs: if a pig is transported over more than 100 kilometers, the chance of dying during transportation is significant (1%). Capital loss of a pig is about € 125 per load. If pigs have to be transported more than 200 kilometers, a penalty cost of € 100 per pig is added, making this kind of transport very unlikely. The Eco-label organization decided to select 14 slaughterhouses as suitable for the Eco-label chain.

The Eco-label pork is sold via a channel of selected butchers. This organization of selected butchers recommends about 13 wholesalers to use. These wholesalers are the set of possible locations in the model with given capacity. Transportation costs are calculated based on the duration of the transport, defined as load and unload, outbound journey and return journey. Load and unload take about 90 minutes (€50 per hour), the load unit is about 150 carcasses.

4.1.3 Slaughtering costs

Traditionally, farmers sell pigs to the slaughterhouse and get a certain price per kilogram in return. Costs of slaughtering are already part of this price. A different structure is valid for Eco-label pigs. The Eco-label organization buys the approved pigs back from the slaughterhouse to sell them to the wholesalers. Therefore, only the slaughtering costs have to be taken into account. Because all other costs are calculated in number of pigs, an average amount of kilograms per pig is assumed. Slaughtering costs show economies of scale. The more pigs to slaughter in a year, the less the slaughter costs per pig are. We assume these costs are the same for

---

[4] http://www.cbs.nl

each slaughterhouse. In practice, the calculation has many graduations. A different cost price can be calculated for each possible interval. This is hard to model in a mixed integer linear programming model. Therefore this situation is simulated by using a piecewise linear cost function (see Figure 3).

The variable costs are as follows: the first 52 000 pigs are slaughtered for € 11.5 per pig, the price drops to € 9,50 for anything between 52 000 and 130 000 pigs, and anything in excess of 130 000 pigs costs € 6,50 per pig. The variable costs for a load of 100 000 pigs are e.g. 52000 * 11.50 + 48000 * 9.50 = € 1 054 000 (€ 10,54 per pig). Cost reductions mentioned above are higher than the cost reductions in practice to compensate for the fact that only a small number of segments are chosen in the cost function, with a fixed price in each segment.

Fixed costs are 0 if the slaughterhouse is not used for slaughtering Eco-label pigs. Note that in traditional location analysis fixed costs arise for opening a facility. In the Eco-label problem we assume that the facility is used anyway by traditional pigs, so the decision to use a slaughterhouse for Eco-label pigs has no consequence for opening or closing a facility. The same argument holds for the wholesalers. As the wholesalers already exist, it is quite easy to change from one wholesaler to another, without consequences for opening or closing a facility. This is only true if the share of Eco-label pigs is quite small compared to the number of traditional pigs.

Schrage (1991) describes how this piecewise linear cost function can be represented. Three segments can be distinguished. The first segment is between 0 and 52000, the second segment is between 52000 and 130 000, and the third segment is anything above 130 000. Let c_1, c_2, and c_3 be the slopes of the segments. The first intercept $K_0 = 0$. This means, no costs arise if the slaughterhouse is not used. The intercepts K_1 and K_2 are determined from the formula $K_{i+1} = K_i + c_i v_i - c_{i+1} v_i$. V_i is defined as the level of pigs where costs have a breakpoint. The cost curve represents an activity with economies of scale, therefore it is concave, piece-wise linear and continuous for positive outflow, so $c_{i+1} < c_i$ (see Krarup and Pruzan, 1983).

The parameters can be now computed as:

$c_1 = 11.5$	$v_0 = 0$
$c_2 = 9.5$	$v_1 = 52000$
$c_3 = 6.5$	$v_2 = 130000$
	$v_3 = \infty$

$K_1 = 0 + 11.5 * 52\,000 - 9.5 * 52\,000 = 104\,000$

$K_2 = 104\,000 + 9.5 * 130\,000 - 6.5 * 130\,000 = 494\,000$

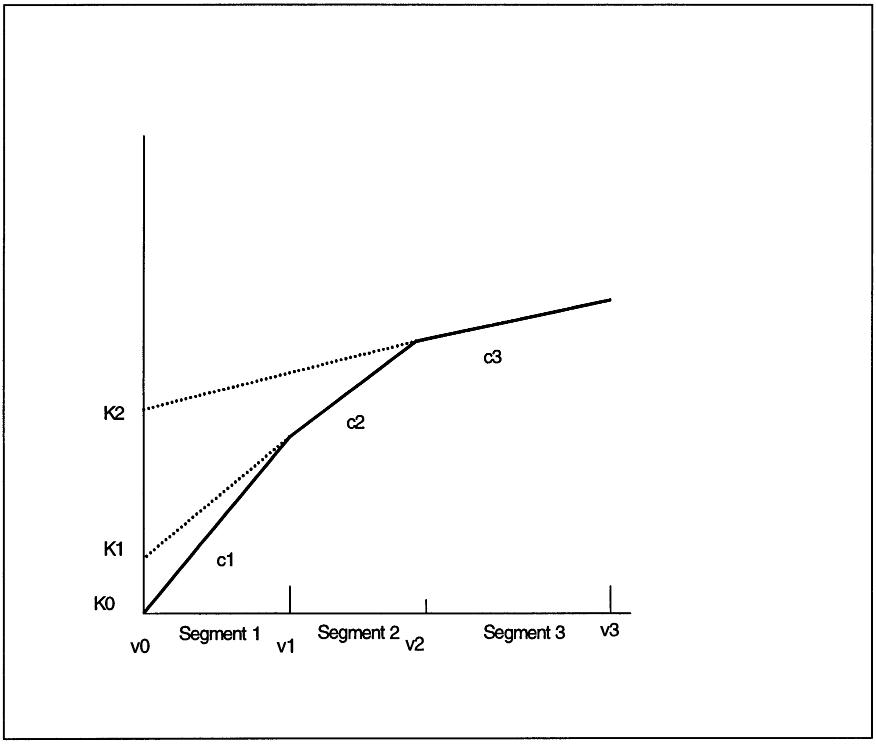


Fig. 3. Piecewise linear cost function

4.1.4 Multiple levels and multiple periods

Because of the lack of investment costs (the facilities already exist so do not have to be built and are used anyway by the traditional pigs husbandry), some modifications compared to the standard two-level multi-period capacitated facility location problem can be made.

First, the fixed costs for opening a facility are 0 in this special case. If the variable costs are already linear, which is the case for the wholesalers, it is not necessary to define a Boolean variable that is 1 if the facility is used, and 0 if not. However, we did use a Boolean variable for generality purposes. As mentioned in the previous paragraph, the slaughterhouses do need a Boolean variable because of the structure of the cost function.

Second, it is rather easy to switch between facilities. No extra costs arise from changing from one wholesaler in period 1 to another wholesaler in period 2. These circumstances are different from traditional cases in the literature. It is therefore possible to decompose the multi-period model into independent static single-period problems for each of the five periods. We assume that the only parameter changing over time is the production of Eco-label pigs in the farm areas. However, if it appears to be difficult to change between wholesalers or slaughterhouses because of long-term contracts, the problem has to be solved as a dynamic multi-period problem, see e.g. Gue (2003).

4.2 Problem Formulation

4.2.1 Notation

The following notation is used:

Index sets:

$i \in I$ = *set of producing farm areas*
$j \in J$ = *set of slaughterhouses*
$k \in K$ = *set of wholesalers*
$l \in L$ = *set of breakpoints in the piecewise linear cost curve (see section 4.1.3)*
$t \in T$ = *set of years in the time horizon*

Parameters:

p_{it} = *production of pigs in area* i *(pigs) in time period* t *(years)*
v_{lj} = *number of pigs at breakpoint* l *at slaughterhouse* j (pigs), $v_{lj} > v_{l-1,j}$ $\forall$ l=1,2,3,...
a_j = *maximum capacity at slaughterhouse* j *(pigs)*
b_k = *maximum capacity of wholesaler* k *(carcasses)*
c_{ij} = *unit transportation costs between area* i *and slaughterhouse* j *(€/pig)*
d_{jk} = *unit transportation costs from slaughterhouse* j *to wholesaler* k *(€/carcass)*
e_{lj} = *variable slaughtering costs for breakpoint* l *at slaughterhouse* j *(€)*
m_{lj} = *fixed slaughtering costs for breakpoint* l *at slaughterhouse* j *(€)*
n_k = *fixed costs at wholesaler* k (€)
s_{ij} = *transportation distance from area* i *to slaughterhouse* j *(km)*
t_{jk} = *transportation time per shipment including load and unload from slaughterhouse* j *to wholesaler* k *based on a unit load of 150 carcasses (minutes)*
h = *costs for transporting carcasses (€/minute)*

Note that

1. $c_{ij} =$
$1.363 + 0.0056 * s_{ij}$ if $s_{ij} \leq 100$
$1.363 + 0.0056 * s_{ij} + 1.25$ if $100 < s_{ij} \leq 200$
$1.363 + 0.0056 * s_{ij} + 100$ if $200 < s_{ij}$

2. $d_{jk} = 2 * t_{jk}/150 * h$, *because transportation costs have to be assigned to both the forward and return journey;*

3. $m_{l+1,j} = m_{lj} + e_{lj} * v_{lj} - e_{l+1,j} * v_{lj}$ $\quad l \in L, ,j \in J$

Decision variables:

W_{ijt} = *number of pigs transported from area* i *to slaughterhouse* j *(continuous) in period* t
X_{jkt} = *number of carcasses transported from slaughterhouse* j *to wholesaler* k *in period* t
U_{ljt} = *number of pigs in slaughterhouse* j *in period* t *if between* $v_{l-1,j}$ *and* v_{lj} *pigs are slaughtered, 0 otherwise (semi-continuous)*

Y_{ljt} = *1 if* $v_{l-1,j} < U_{ljt} \leq v_{lj}$ *in period* t, *0 otherwise (bilinear).*
Z_{kt} = *1 if wholesaler* k *is used in period* t, *0 otherwise (bilinear)*

4.2.2 Model formulation

In this case, the Eco-label Location problem (ELP) can be decomposed in t independent static one-period problems ($ELP_{t)}$, with z(ELP) = Σ_t z(ELP_t). Total costs of the location and assignment choice over five years can be defined as the sum of the total costs of the location and assignment choice for each year individually. The formulation of ($ELP_{t)}$ is as follows:

$$Z(ELP_t) = \min \Sigma_i \Sigma_j \; c_{ij} W_{ijt} + \Sigma_j \Sigma_k g_{jk} X_{jkt} + \Sigma_j \Sigma_l (m_{lj} \; Y_{ljt} + e_{lj} U_{ljt}) + \Sigma_k n_k Z_{kt} \quad (2)$$

subject to:

$$\Sigma_j W_{ijt} \geq p_{it} \qquad \forall i \quad (3)$$

$$\Sigma_i W_{ijt} \leq a_j \Sigma_l Y_{ljt} \qquad \forall j \quad (4)$$

$$\Sigma_i W_{ijt} = \Sigma_l U_{ljt} \qquad \forall j \quad (5)$$

$$\Sigma_l U_{ljt} = \Sigma_j X_{jkt} \qquad \forall j \quad (6)$$

$$\Sigma_j X_{jkt} \leq b_k Z_{kt} \qquad \forall k \quad (7)$$

$$W_{ijt}, X_{jkt} \geq 0 \qquad \forall i,j,k \quad (8)$$

$$Y_{ljt}, Z_{kt} \in (0,1) \qquad \forall j,k,l \quad (9)$$

The objective (2) is to minimize the sum of total transportation costs (inclusive penalty costs) and total fixed costs and variable costs of the slaughterhouses and total fixed costs of the wholesalers in year *t*. Note that in this special case fixed costs for the wholesalers are 0 (see section 4.1.4). Constraints (3) state that all production in the farm areas must be transported to a slaughterhouse. These constraints reflect the push character of the model, i.e. all produced Eco-label pigs will be sold by the butchers due to the Eco-label covenant between the supply chain members. Constraints (4) state that total flow to slaughterhouse *j* can never exceed the capacity of slaughterhouse *j*. Constraints (5) are flow balancing constraints, stating that all pigs transported to slaughterhouse *j* will be slaughtered there. Constraints (6) are flow balancing constraints, stating that all pigs slaughtered in house *j* will be transported to the wholesalers. Constraints (7) state that total flow to wholesaler *k* can never exceed the capacity of wholesaler *k*. Finally, constraints (8) are the non-negativity constraints on the flows, and constraints (9) are the binary constraints on the location variables.

4.2.3 Solution method

For our test case, the problem dimension of (ELP_t), (i,j,k,l), is (28, 14, 13, 3), i.e. 28 farm areas, 14 potential slaughterhouses, 13 potential wholesalers and 3 breakpoints in the piecewise linear cost curve. This results in a model with 671 decision variables (574 continuous, 42 semi-continuous, and 55 binary variables) and 84 constraints. The one-period mixed integer linear program of this size can still be solved to optimality by exact methods like branch and bound. We used GAMS 19.8 (GAMS Integrated Development Environment, 2001) on a PC to solve the problem. For this problem size, the computation time was negligible. If the problem cannot be solved for each period individually, the problem size will increase to (i,j,k,l,t)=(28,14,13,3,5), which results in about 3000 continuous, 210 semi-continuous and 275 binary variables. Also, the number of constraints will increase to at least 400 constraints except for extra conditions modelling the relation between the solution of one period and the solution in the next period. If this is the case, it is not sure that the problem can still be solved by general branch and bound software in a reasonable time.

5 Analysis of Scenarios

We analyzed several scenarios to find an efficient distribution network of Eco-label pigs in the long term. The Eco-label organization expects a growth up to 1 million Eco-label pigs in year 5 (grow rate 2.07), given that each area grows with the same rate. The actual number of Eco-label pigs of about 26000 was measured in the year 2001 (year 0). Therefore, the scenario is carried out for the period 2002-2006 (year 1 – year 5). This is the base-scenario (S0). To analyse the sensitivity of this assumption, we also analysed a scenario (S1) with a growth up to 500000 Eco-label pigs in year 5 (grow rate 1.8). Further, we analysed the scenario that growth takes place only in the South part of the Netherlands, both up to 1 million (S2) and 500000 pigs (S3) because of the current situation. The Eco-label organization has more business relations in the South of the Netherlands than in other areas of the Netherlands, not just with the farmers but also with the food producers, who are of great influence on the farmers. So, the chance of convincing a farmer to join the Eco-label is much more likely to succeed in that region.

Furthermore, we investigated the effect of the construction of the slaughtering costs on the results by varying the cost parameters c_1, c_2, and c_3 (see section 4.1.3) from (11.5 / 9.5 / 6.5) to (11.5 / 10.5 / 9.5), (S0A) and (11.5 / 11 / 10.5), (S0B). In total, 6 different scenarios were analysed. For reasons of readability, we only report extensively on the base scenario (S0), and discuss the results of the other scenarios more briefly. Results of the base case scenario analysis can be found in Table 1.

The results show the use of more and more slaughtering houses in the later years, with higher production. In year 1, only one slaughtering house is used in the optimal solution, while in year 5 seven slaughtering-houses are used. The same holds for the wholesalers: one or two in the first years to ten in year 5.

Table 1. Used slaughtering-houses (H) and wholesalers (W) in the optimal solution for the base case scenario for a 5-years period.

H1	H2	H3	H4	H5	H6	H7	H8	H9		Total
					100%					54333
								100%		112538
	100%									233095
				27%	46%		27%			482800
13%	22%	13%	13%	13%	13%	13%				1000000
W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	Total
				52%				48%		54333
	100%									112538
	100%									233095
				68%	5%	11%	11%	5%		482800
10%	25%	8%	5%	34%	3%	5%	5%	3%	3%	1000000

Scenario (S1) with a growth to half a million Eco-label pigs has great resembles with the base case scenario. One slaughtering-house is used in the optimal solution for year1 up to year 4, only in year 5 three slaughtering-houses are used. The formulation of the slaughtering costs has a high impact on the number of slaughtering houses used. It is optimal to fill just one slaughtering-house up to the breakpoint with the largest financial advantage and the largest economy of scale (if capacity allows). More slaughtering-houses are only used if total production is either too small to fill a slaughtering-house to this breakpoint, or the production is large enough to fill more slaughtering-houses up to the breakpoint. This rule of thumb is only disturbed if the penalty costs for long distances become too high.

Results of scenarios (S2) and (S3), growth only in the south of The Netherlands, show almost always an optimal solution with only one slaughtering-house and one wholesaler each year. This can be explained by the small surface of the area of production. The transportation costs have a high impact on the choice which slaughtering-house and wholesaler to use. Transportation costs from the farm areas to the slaughtering-houses are much larger than the transportation costs between slaughtering-houses and wholesalers. This cost difference is caused by two reasons. First, a wholesaler can often be found near a slaughtering-house and second, animal transportation is more expensive then carcass transportation.

Results from scenario (S0A) and (S0B), on the effect of the structure of the slaughtering costs, show that if the variable cost differences between each segment in the cost function are smaller, the economies of scale are also smaller. In some cases, the economies of scale of the largest segment (more than 130 000) cannot longer compensate higher transportation costs due to long distances.

Summarizing, some potential slaughtering-houses (5) and wholesalers (2) are never optimal. These locations are not interesting to investigate. One slaughtering-house (H6) appears more than proportional in the optimal solution over all scenarios. About three other slaughtering-houses appear regularly in the optimal solutions, the remaining slaughtering-houses appear only once in a while in the optimal solution. In most occasions, only one or two wholesalers are chosen in the optimal solution. Which wholesaler to use depends on which slaughtering-house is used. It appears that optimal combinations exist of slaughtering-houses and wholesalers.

6 Conclusions

An efficient distribution network for Eco-label pigs can be found by solving a (multi-period) two-level capacitated facility location problem. The size of the problem is such that standard mixed integer programming software can solve the problems in very reasonable time, so many scenarios can be analyzed.

In the actual situation, only one slaughtering-house and one wholesaler are part of the Eco-label pigs supply chain. It appears from the results that this situation can hold for the first years to come. Only if the production of Eco-label pigs really increases, more locations are needed. The Eco-label organization prefers a small number of facilities to cope with, because of quality checks and certifications. However, it is remarkable that the optimal locations for slaughtering-houses and wholesalers do change regularly over the years. Therefore, we can advise the Eco-label organization not to stick with one slaughtering-house or wholesaler but to keep an open eye to alternative facilities.

With respect to the distribution network, we can conclude that the structure of the slaughtering costs has a high impact on the choice how many slaughtering-houses to use, whereas the transportation costs influence the choice, which slaughtering-house to use. Some potential slaughtering-houses are never chosen. Taking all scenarios into account, we advise two slaughtering-house locations to use because they are efficient during the total of five years of investigation. The optimal wholesaler locations depend on the used slaughterhouses.

The approach presented in this paper can be used for location problems with the following characteristics:

- Discrete multi-level multi-period capacitated facility location problems
- Living products or perishable products (no inventory)
- Quick growth of a new brand in an existing market
- Facilities are already available due to the existing market, so almost no investment costs are required.

Areas of interest to think about are e.g. Eco-label Fruit, Vegetables or Flowers but also green brands in gasoline (Shell Pura), or green electricity. Another area of interest is the growing segment of reused or remanufactured products that are sold as a B-grade flow parallel to the traditional product flows.

Acknowledgement

The authors thank the management team of De Hoeve b.v. for providing the interesting problem description and their contribution to the research and the Foundation for Agricultural Chain Knowledge (AKK) for financing this research.

References

Barros, A. I. / Labbé, M. (1994): A general model for the uncapacitated facility and depot location problem, Location Science 3, 173-191.

Bloemhof, J. M. / Salomon, M. / Van Wassenhove, L. N. (1996): The capacitated distribution and waste disposal problem, *European Journal of Operational Research*, Vol. 88, 490-403.

Blonk, T. J. (2001): LCA varkensvleesproductie, Blonk Milieu Advies, Gouda, NL.

Brandeau, M. L. / Chiu, S. S. (1989): An overview of representative problems in location research, *Management Science*, Vol. 35, 645-674.

Canel, C. / Sidhartha, R. (1999): The uncapacitated multi-period facilities location problem with profit maximization, *International Journal of Physical Distribution & Logistics management*, Vol. 29 (6), 409-433.

Chopra, S. / Meindl, P. (2001): Supply Chain Management, Prentice Hall, Inc., Upper Sadle river, New Jersey.

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*, Vol. 50, 280-297.

Francis, R. L. / McGinnis, L. F. / White, J. A. (1983): Locational analysis, *European Journal of Operational Research*, Vol. 12, 220-252.

Gue, K. R. (2003): A dynamic distribution model for combat logistics, *Computers & Operations Research*, Vol. 30, 367-381.

Harkness, J. / Revelle, C. (2003): Facility location with increasing production costs, *European Journal of Operational Research*, Vol. 125, 1-13.

Heizer, J. / Render, B. (2001): Principles of operations management, Prentice Hall, Inc., Upper sadle river, New Jersey, 308-309.

Krarup, J. / Pruzan, P. M. (1983): The simple plant location problem: survey and synthesis, *European Journal of Operational Research*, Vol. 12, 36-81.

O'Kelly, M.E. (1995): Inferred ideal weights for multiple facilities, in: Drezner, Z. (eds.): Facility location, Springer, New York, 99-102.

Plastria, F. (1995): Continuous location problems, in: Drezner, Z. (eds.): Facility location, Springer, New York, 225-262.

Schrage, L (1991): LINDO, an optimization modeling system, 4th edition, The scientific Press, USA.

Sule, D.R. (2001): Logistics of facility location and allocation, Marcel Dekker, New York.

Tcha, D. / Lee, B. (1984): A branch-and-bound algorithm for the multi-level uncapacitated facility location problem, *European Journal of Operational Research*, Vol. 18, 35-43.

Online Shopping and Changes in Mobility

Joachim R. Daduna[1] and Barbara Lenz[2]

[1] University of Applied Business Administration at Berlin, Berlin, Germany
[2] German Aerospace Center / Institute of Transport Research, Berlin, Germany

Abstract Specific information on the relations between current and forecast online shopping developments and changes in commercial and private mobility structures is scarce. A number of assumptions exist, however, most of which are not backed by sufficient empirical data. After explicating the considerations underlying the substitution thesis, the complementarity thesis and the induction thesis, this article investigates, based on various fields of online shopping application, whether and in what form these theses can actually be relevant. The fields of application include (industrial) direct marketing, the mail-order selling (in its various expressions), and the food / non-food retail trade. The (scant) empirical findings available are moreover reconciled with the three theses. Finally, the effects that could arise for logistic services are investigated, particularly with a view to new structures and new service offers.

1 Online Shopping and Transport: An Estimate of Potential Effects

In recent years, changes have occurred in the retail trade with the rise of online shopping. The future effects of these changes are assessed in highly different fashions. This relates to the scope of the development of customer potential on the one hand and to changes in mobility structures on the other. The realization of customer potential is strongly influenced by technical factors (availability of Internet access, familiarity in dealing with software tools) and the suitability of the products (cf., e.g., Daduna 2003). The mobility effects, in contrast, must be viewed as a factor directly dependent on the potential realized; i.e. it is a development largely determined by external influences.

The conclusions drawn by any analysis of the effects of online shopping on mobility structures depend not least on the definitions made. The estimation and study of the transport-relevant effects of online shopping are normally based on two different types of meanings. A narrowly drawn understanding of the relation between online shopping and transport is sometimes assumed, which includes only activities directly related to a purchase transaction; i.e. the subject of study is limited to the "purchase made," whether physical or virtual. The second meaning is more broadly defined, involving information procured online, i.e. information which later leads to a (selective) purchase transaction in the stationary retail trade (cf., e.g., Daduna 2003). This more comprehensive definition must be assumed if the goal is to capture *all* potential, mobility-related effects, because, depending on

B. Fleischmann et al. (eds.), *Distribution Logistics*

the products, purchase decisions are often based on complex information and decision-making processes.

The mobility structures and transport to be taken into account in connection with online shopping primarily relate to private and commercial travel incurred or eliminated in connection with (logistical) activities in the *business-to-consumer-*(B2C)-*segment* between dealers or trading companies and (end)consumers, though transportation in the *business-to-business*(B2B)-*segment* within (commercial) distribution structures can also be affected. When considering potential transport effects in the B2B-segment, a basic distinction must be made between shopping traffic on the demand side and delivery transport on the supply side (cf. also Klaus *et al* 2002). While consumer shopping traffic can be substituted, supplemented through add-on effects or additionally induced, transport volume arises in all cases on the supplier side, because additional delivery services have to be rendered. Just how comprehensive this additional transport can be is evident from the organization of the delivery services, which in turn mainly depends on the possibility for integration into previously existing service structures. The net effects which therefore result as a whole from online shopping depend both on consumer behavior and on the organization of distribution logistics.

The key items discussed in this paper relate to the traffic effects based on changing consumer shopping behavior and to the related potential effects on the part of suppliers, who now have to organize the delivery of the goods. The point of departure for this discussion will be three different theses, which assume sharply different effects, but primarily relate to the demand side (cf., e.g., Lenz 2003a). An estimate of effects more in line with reality will thus arise when effects on suppliers are also taken into consideration.

- ***Substitution thesis*:**

  According to this thesis, savings are achieved through online shopping in the B2C-segment in the form of the travel volume expended by consumers. The savings result from the fact that the need for a physical purchase on the part of consumers is eliminated due to the online purchasing. This means on the one hand that the number of purchases is reduced, so that consumers travel less frequently for shopping. On the other hand, previously combined shopping transactions are broken apart, thus shortening the travel chain.

  In the best case, the substitution effects on the consumer side are not overcompensated by the delivery process which has to be carried out by the supplier because economies of scale conditioned by efficient delivery services come into play. The net effect of online shopping then in fact consists in the substitution of traffic. However, it must be additionally investigated at this point what changes occur in the B2B-segment, because the logistical processes between producers and the retail trade sometimes have to be restructured through the addition of services in the „last mile".

- ***Complementarity thesis*:**

  Here it is assumed that shopping trips will continue to be incurred to an unreduced level, even though parts of the demand can be covered via online shopping. This means that though consumer shopping behavior changes through online shopping, their shopping travel volume does not.

Complementarity on the demand side is in all cases associated with additional transport volume arising as delivery services. An increase in traffic thus arises overall. Secondary consequences resulting from the separation of shopping activities are not taken into account here. The separation may lead to changes in the structure of retail locations, in particular to increased decentralization, and likely related changes in consumers' choices of transport means.

- ***Induction thesis*:**

  Here an increase in shopping-related traffic volume is basically assumed on the part of customers. This increase can come about because the Internet offers customers access to more information about retail locations with which they were not yet familiar, but which they start to use, thus generating more traffic. The increase in traffic can also occur, however, because the time saved through online shopping can be used for additional mobility-generating activities, leading on balance to an increase in traffic volumes. To the additional traffic arising on the part of consumers can be added the delivery transport made necessary through online shopping.

In summary, it can be concluded at this point that the possibilities for traffic reduction through online shopping have to be investigated in detail. The probability of a reduction in traffic volume is only high if consumer behavior changes, while "intelligent" delivery systems support this behavioral change in favor of a traffic volume reduction.

No uniform view can be identified when analyzing the past studies concerning online shopping's effects on mobility structures. This is due *inter alia* to the complexity of the relations between effects, which cannot be measured at all or can only be measured to a limited degree. This pertains not only to shopping itself, but to shopping travel behavior. Only limited information exists at any rate concerning the spatial shopping behavior of consumers. In particular, a basis for comparing online shopping to shopping in traditional commercial structures, which is sufficiently quantified to be able to identify changes, is missing. Furthermore, the varying market / product segments must be considered in a differentiated fashion, because online shopping routines in the food / non-food retail trade, for instance, will certainly differ from those related to media products.

Therefore, an approach will be developed below that discusses the potential effects of online shopping on transport considering both consumer behaviour and commercial distribution. The basis behind this approach is the idea that relevant effects are mainly due to the following factors :

- The acceptance of products by consumers and thus the market segments in which online shopping promises to take on greater volume.
- The product-specific shopping travel behavior of consumers.
- The selling concepts / marketing channels through which these products have reached consumers to date.

Finally, the effects that can arise through changes in customer shopping behavior on processes for the creation of services among logistics service providers will be shown, particularly with a view to new structures and innovative services. Empirical data, however, cannot be provided for all aspects because relevant studies are missing on the national as well as on the international level.

2 Online Suitability and Market Segmentation

Methods for determining the suitability of products for online shopping frequently stem from marketing research work. Market research normally divides the (stationary) retail trade into three segments (c.f., e.g., Holton 1958; Bucklin 1963; Halder et al. 1999):

- *Convenience goods* in the sense of "daily necessities", e.g. food, drugstore goods.
- *Shopping goods* in the sense of "occasional goods", e.g. clothing and shoes.
- *Specialized goods* in the sense of normally high-quality, long-lasting goods, such as furniture and large devices in the brown and white merchandise categories.

Methods for estimating transport effects apply the above division to online shopping (cf., e.g., Janz 2001). Other studies differentiate goods much more sharply by product groups (cf., e.g., Luley et al. 2002). However, because differences arise within these product segments with respect to the purchase decisions associated with individual products, a simplification is made in these approaches and the bases for estimating the effects generated in this fashion are generalized to a certain degree, something which is thoroughly comprehensible in empirical studies for reasons of operationalization (cf., e.g., Merkel / Kromer 2001; Merz 2002: 587).

Alternatively or supplementarily, whether products are suited for online shopping can be determined by differentiating among the purchase decision making processes (cf., e.g., Daduna 2003). Assuming a relative interpersonal similarity of customer behavior[1] when it comes to purchase decision-making processes, the primary question that arises is whether an *individual* decision is made (and thus a decision directly related to a *particular* product) or whether a selected product is merely picked out by coincidence during the purchase from a number of *completely identical* and thus *exchangeable* product units. For instance, it can be assumed in the food / non-food retail trade that fresh goods (fish, meat, vegetables, fruit, etc.) lack any clear *online affinity* because customers in this segment normally reach decisions based on the specific case. This is also true for parts of the clothing market segment, e.g. women's and men's outerwear, where a "look-&-feel" effect in the sense of an *event purchase* must be taken into account (cf., e.g.,

---

[1] This similarity in shopping behavior is in no way self-evident. Differences in behavior between social-demographic groups can be documented both for physical and for online shopping (cf., e.g., Lenz 2003).

Bretzke 1999; Merkel / Kromer 2001; Dach 2000). This "look-&-feel" effect simultaneously helps the customer to reduce the purchase risk he or she subjectively feels (cf., e.g., Bretzke 1999). In contrast, an online affinity can be seen in "anonymous" or "non-specific" products, where the customer selects a certain product, without explicitly specifying the product unit in the decision-making process, as in the food / non-food retail trade with standardized products such as beverages, preserves, detergents, etc. and / or with books, (technical) household appliances, textiles (for daily use), etc. in the mail-order selling.

The fact that shares of "anonymous" or "non-specific" products can be found in all product groups is confirmed by the fact that consumers who today already practice online shopping or are open to the idea of shopping online in the future indeed assume an increase in electronic shopping in all product groups. A differentiated study conducted in the Stuttgart region (Germany) confirms corresponding growth expectations on the part of consumers (Figure 1).

The statements on the acceptance-conditioned online suitability of products and the related market segmentation emphatically make clear that any analysis of potential changes in consumer mobility behavior must be differentiated to a high degree. Additional factors, which must likewise be taken into account in any analysis, relate to the *coupling effects* arising during individual mobility activities (cf., e.g., Frank / Frechen 1998; Burgdorff 2000, p. 20; Mindali / Salomon 2003), which impede or even sometimes prevent an exact allocation of individual activities. The question of whether and to what extent consumers behave rationally by

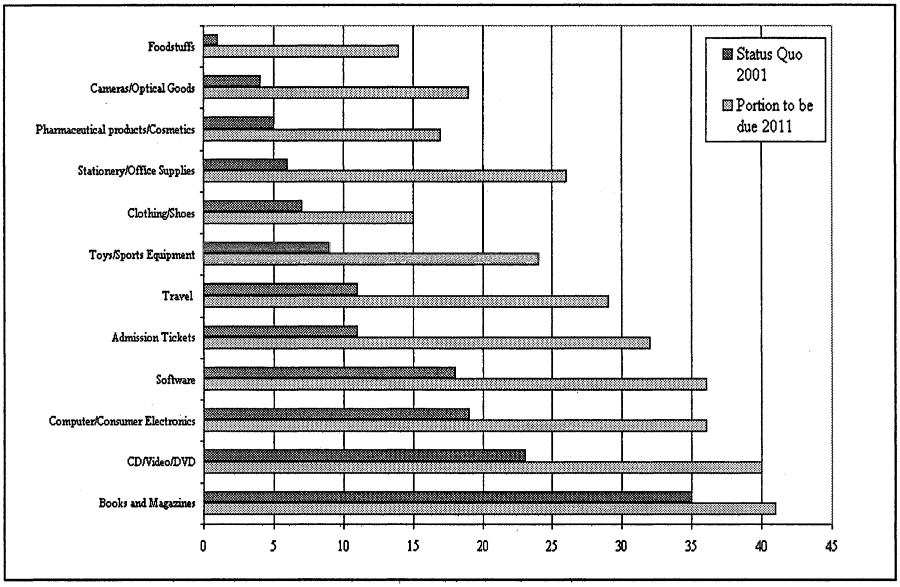


Fig. 1. Estimate of future shopping behavior by E-commerce-friendly Internet users* (source: Luley et al. 2002)
(* Consumers who already purchase online or plan to do so in the near future and have no reservations about online shopping are considered "e-commerce-friendly".)

explicitly including the related *purchasing costs* (e.g. travel costs, etc.) in their individual decisions is also worthy of fundamental consideration. Analogously, this also holds true for the *information procurement* activities accepted when making decisions; these play a major role, particularly with high-quality products. The question of the routines related to shopping behavior or shopping travel behavior that basically stand in the way of change also arises in this context.

3 Online Suitability in Different Marketing Channel Structures

In addition to the suitability of products for online shopping, substantial traffic-related effects can be expected from the type of marketing channels used for products: The greater the similarity of the current marketing and distribution to online shopping, the less the potential effects. Three segments can be differentiated in relation to the marketing channels, the basic structures of which must be distinguished: the direct marketing conducted by manufacturers, the (classical) mail-order selling including the (exclusively) online-based limited-line retailer, as well as the food / non-food retail trade (cf., e.g., also de Koster 2002; Daduna 2003).

- ***Direct marketing through (industrial) manufacturers***

  In this form of selling, classical structures (e.g. the implementation of sales organizations) are largely foregone so that communication and the conclusion of contracts occur directly between manufacturers and consumers. These marketing structures are of particular interest in relation to models for *customer configured manufacturing*. Examples include companies that orient their offer and manufacturing structures on *built-to-order concepts* and / or *mass customization approaches* (cf., e.g., Reichwald / Piller 2000). The aim is to satisfy the increasing customer-specific demands through consistent *user-specific* product configuration. This, however, requires a seamless path of communication between manufacturers and customers, as offered by the Internet.

  The most well-known example of consistent implementation of a built-to-order approach is the company *Dell Computer* (cf., e.g., Karg 1998), which largely produces on the basis of customer orders. This highly successful business strategy has made the company one of the world's leading PC manufacturers. Another example of the approach, this time from the clothing industry, consists in the generation of (technical and user-specific) *product variations* using mass customization technology (e.g. tailor-made shirts). In this way, the uniformity of some products can be reduced by enabling customers to help design the products to a certain degree. The focus in the direct marketing of customer configured manufacturing currently lies in products from the computer sector (cf., e.g., Merz 2002, p 56). Considerations in the automotive industry - which no longer necessarily have to be seen as visionary - assume that customers will be able to configure their cars virtually in the foreseeable future (cf., e.g., Hirn 2002). To what extent this possibility will actually be implemented is an open issue, however, given that the *effect of experiencing* *the product live*, e.g. in a (*non-virtual*) test ride, can have a major influence on the purchase decision.

- ***Mail-order selling***

In this area, the existing segments must be differentiated, the difference lying in the scope and structural form of the e-commerce activities. On the one hand, there is the (classical) mail-order trade, where online shopping is an additional marketing channel through which existing structures can be deliberately complemented (and sometimes even substantially extended) (cf., e.g., Bretzke 2000). On the other, there is a largely specialized trade, which uses online shopping as its exclusive form of marketing.

(Classical) Mail-order selling

The classical mail-order selling is characterized by an assortment which is essentially equivalent to that of a department store in terms of breadth and depth (e.g. the international mail-order companies Otto Versand and Quelle AG). However, there are also suppliers with a more narrow assortment, which are comparable in terms of assortments with limited-line retailer (e.g. Bader-Grossversand GmbH). These are normally companies that have been on the market for a long time. Online shopping merely represents for them an extension - even if a highly attractive one - of existing marketing channel structures, thereby enabling them to capture new customer potential (cf., e.g., Palombo / Theobald 2000). One distinguishing feature in this market is currently the (still) very pronounced dominance of classical communication instruments (such as printed media), though online commerce is making clear gains.

Limited-line retailer focusing on online shopping

This segment involves wholesale trading companies, which orient their communications and thus also their marketing channels in a consistently mono-structured fashion by presenting themselves exclusively on electronic markets. Amazon.com, Inc., the book and media trading company, is one example of such a business (cf. also Wirtz 2001, pp 235.). These companies too often have only been on the market for a few years and normally settle their sales exclusively via online shopping. In addition to the computer industry, which was already mentioned with respect to direct marketing, the focus here is mainly on media products (cf., e.g., Merz 2002, p 56).

- ***Food / non-food retail trade***

For the stationary shopping outlets in the food / non-food retail trade (regarding the types of retail businesses, cf., e.g., Tietz 1993, pp 30; Müller-Hagedorn 1998, pp 41), it can be assumed that the entire assortment of a supermarket will not be suited to online shopping. With certain types of merchandise (e.g. fresh goods), customers want to select themselves the "piece" which they actually buy. These products can thus not be said to have an online affinity, given that a customer will normally not agree to an "allocation" over which he or she has no control. Moreover, the technical (transport) problems, which are associated with additional costs, must also be considered (cf., e.g., Kuchenbecker / Siebel 2001). The question of the achievable long-term profitability (and acceptance) necessarily arises at this point. The situation is different with respect to "non-specific" (standard) products that can be offered in a closed package or packing (beverages, preserves, detergents, etc.) and have a constant (and even guaranteed) quality and are usually in regular demand (inventory purchases with a high frequency of repetition). Such products can be viewed as having a strong online affinity, as the decision to buy is based on the expectation of homogeneous quality.

It cannot be expected that online shopping will fundamentally change the existing structures of this retail trading segment in the foreseeable future, though it will develop into an additional form of selling. This will have substantial structural effects, for the existing supply chain must be extended or modified by suppliers, given that a "shift" towards individual customers is necessary. Schögel et al. (1999) thus speak in this context of an *extended front-end*. The advantages of online shopping on the supplier side must largely be viewed in light of long-term goals, the achievement of which will essentially depend on the future market penetration of this form of selling. On the buyer side, the (physical) relief from having to make routine purchases can be a major benefit, though a series of factors must be taken into account. Decisive will be whether customers accept the necessary separation in the (individual) shopping transactions resulting from the merely partial online affinity in this market segment.

Suppliers such as *FreshDirect* (http://www.freshdirect.com), which deliberately carry a fresh goods assortment, act at variance with these basic structures. However, their products are not stored, but are procured, processed and delivered at short notice in accordance with customer orders. Given that location-bound structures are concerned here, too, as a result of framework conditions, we will not deal with this form of commerce in the following.

4 Basic Forms of Delivery in Online Shopping

When discussing the traffic-related effects of online shopping, in addition to the logistical routines in the B2B-segment, one must also bear in mind the activities incurred in the „last mile“, which are of decisive importance when estimating transport expense. For this reason, before discussing the individual segments, we will detail the key service structures in the B2C-segment. The main issues in this context are direct delivery to customers and the integration of pick-up points as (spatially) remote interfaces in the necessary delivery services. Even if customers tend to prefer direct delivery (cf., e.g., Wagener et al. 2002, pp 45; Weber et al. 2002, pp 84), pick-up point models must also be taken into account, since they are already of importance in subsegments of the mail-order selling. Before analyzing (potential) market-segment-related changes in mobility structures, we will consider the basic structures of the two approaches in detail below (cf., e.g., Daduna 2003).

- ***Direct delivery***

  With this form of delivery, shipments are delivered directly to the customers (or to a commissioned person) at an agreed location (*home delivery*). Different alternatives are possible regarding delivery conditions, e.g. deliveries at a *point in time* or within a *time window* determined by the customer, on a specified (week)day, after business hours (e.g. between 5:00 and 9:00 p.m.) or on Saturday. Many possibilities exist, though the question of costs plays a decisive role, particularly with express deliveries and agreements on repeated delivery attempts. Of decisive importance for the acceptance of home delivery systems is a realistic comparative calculation that also includes the individual (financial and time-related) expenses associated with purchases in traditional commercial structures (cf., e.g., Punakivi / Saranen 2001).

  Conditioned by direct customer contact, one advantage of direct delivery lies in the simplification of payment, which can be rendered directly when the goods are delivered. Customers may also check the delivery immediately and make any complaints, except in cases where it is necessary to test technical equipment. This normally simplifies return management and, if necessary, the treatment of empties. The main problem, however, is that someone must be present to accept delivery. The question of the distribution and / or allocation of the costs incurred for the delivery service must also be considered.

  The personal acceptance problem associated with direct delivery can be solved to a certain degree through box systems (cf., e.g., Merz 2002, p 449), though the acceptance issue still plays a significant role (cf., e.g., Pflaum 2003; Cairns 2003). The basic idea is to temporally decouple the delivery process from the (physical) acceptance by the customer, whereby the key issue for the delivery company is its flexibility in structuring the delivery services. Installing delivery box systems is not unproblematic, however. Especially in urban areas, it is often impossible to position the systems, which take up a lot of space, near to customers. In addition, it is difficult to finance box systems, given that investments are necessary either through consumers and / or suppliers or an involved logistics services provider.

- ***Integration of pick-up points***

  With this form of delivery, shipments are delivered indirectly to an agreed outlet, a *pick-up point* (cf., e.g., Fichtner / Engelsleben 2001; Merz 2002, pp 449). Provided that there is a complete covering network, shipments can be realized via newspaper stands, cleaners, gas stations and, more generally, shopping outlets with suitable business hours. Fixed outlets with box systems (cf., e.g., Merz 2002, p 449f) and storage containers can also be used, even in drive-in form (cf., e.g., Fichtner / Engelsleben 2001; Merz 2002, p 450). Storage containers and boxes use systems built on the principle of automated small parts warehouses (cf., e.g., Gremm 2001).

For customers, the advantages of intermediary pick-up points lie in the elimination (for the most part) of the temporal commitment to accept delivery. The disadvantage is the additional route, for which costs and the corresponding time worked must be charged. The distance between the location of a pick-up point and that of the consumer is ultimately the decisive issue, likewise for the acceptance of the system. However, this is a procedure from which suppliers can also profit, because they can plan their routing more effectively and reduce the delivery costs through bundling effects.

From the viewpoint of consumers, there is undoubtedly a clear desire for direct delivery with the most exact time specifications possible (cf., e.g., Infas 2001, p 7; Wagener et al. 2002, pp 45), i.e. there are high demands being placed on service and quality (cf., e.g. DVWG 2002). Consumers place considerably less value on the use of pick-up points, but do not rule them out in principle.

5 Effects of Online Shopping on Transport

As mentioned, only limited statements can be made on an empirical basis about the transport effects of online shopping. While studies on the effects on customer shopping traffic provide indications of both substitution and complementary effects, no empirical studies have been conducted to date on the effects on delivery traffic. Assuming the determination made at the beginning that customer and supplier traffic must be considered separately, below we will make a separate evaluation of the effects, which will then be summarized into possible scenarios for transport effects.

5.1 Changes in traffic volume on the supply side

In the following sections the three different segments, direct marketing by industrial manufacturers, mail-order selling, and food / non-food stationary retail trade, are analyzed and discussed from the perspective of suppliers. Particular interest concerns changes in traffic volume caused by online shopping activities.

Direct marketing by industrial manufacturers

Basically, it can be expected that direct marketing by industrial manufacturers will lead to significant changes in distribution structures. Traditional (multi-level) retailing structures are thus being completely eliminated, leading from a logistical point of view to a loss in the bundling effects arising through the integrated warehousing steps and the sales outlets of the retail trade. The (physical) connection between manufacturers and consumers is normally made through *logistics service providers* (courier, express and package services, general cargo shippers, etc.) which assume the transport and sometimes further services (cf. Figure 2).

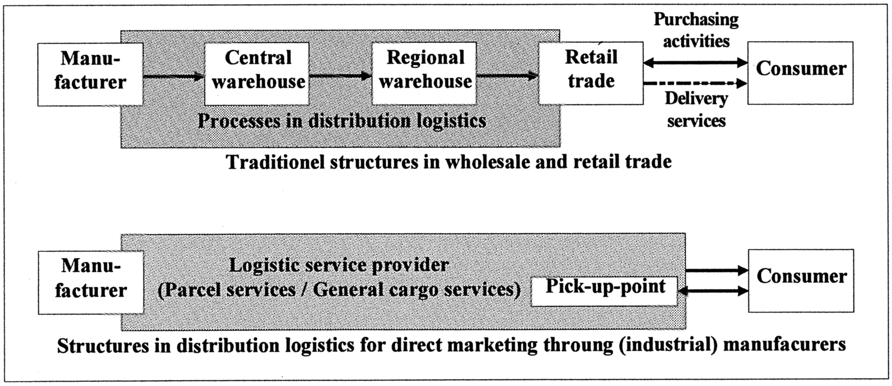


Fig. 2. Basic structures of traditional retail trade and direct selling

Whether and to what extent the elimination of the *bundling effects* on the B2B-level will lead to an increase in transport volume is difficult to predict. It can be assumed, however, that such effects will be fully (or at least largely) compensated under certain conditions through efficient structures in the realm of the logistical service providers, because logistics networks also disclose a high degree of bundling effects. By serving customers in the „last mile", additional (commercial) traffic arises in all cases, though this does not necessarily have to lead to an increase in traffic overall. For example, it must be taken into account that even in traditional commercial structures deliveries took and continue to take place in the „last mile". Moreover, a certain share runs via the pick-up points integrated into the supplier structures (see above). The degree to which a distribution is made between the two delivery forms depends on the framework conditions in each specific case.

Mail-order selling

With mail-order selling, a distinction must be made between the classical structures and suppliers, normally from the limited-line retail trade, who entered the market in recent years and settle sales *exclusively* via *online commerce*. Apart from the development of additional turnover potential, no effects on logistics routines result in the classical mail-order selling (cf. Figure 3). Online shopping here mainly entails changes in communications and contracting procedures, although the customer structure has started to change slightly. In principle, a product-specific differentiation is still made here (package goods, hanging goods, general cargo), which arises from the varying logistical requirements. However, the level of mobility will increase on the consumer level in the package goods segment with the increasing use of pick-up points (cf., e.g., Daduna 2003), leading to a reduction in commercial transport expenses. The question is to what extent this will occur and to what degree this will involve an increase in *motorized individual traffic*. With the other suppliers, who are mostly located in the limited-line retail trade, a different view emerges than in industrial direct marketing.

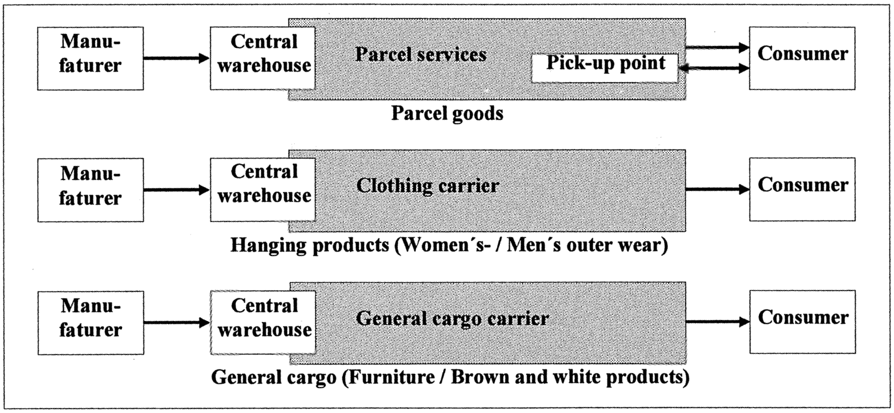


Fig. 3. Logistics structures in the classical mail-order selling

Stationary retail trade in the food / non-food segment

The logistical requirements of online shopping in the food / non-food retail trade are largely characterized by repetitious purchase transactions and deliveries performed at short notice. Relatively narrow effects on transport routines will thus arise on the B2B-level, because multi-stage distribution structures with (decentralized) supply *close to customers* will absolutely be necessary (cf., e.g., Kuchenbecker / Siebel 2001; Cairns 2003). For this reason, potential for a sustained reduction in traffic volumes lies only on the B2C-level (cf. Figure 4).

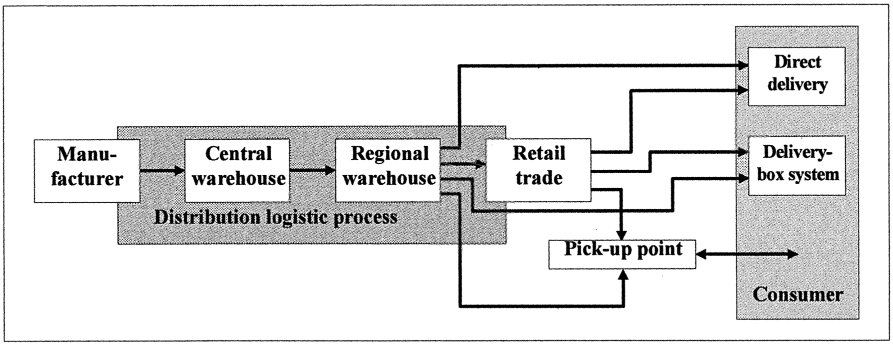


Fig. 4. Basic Structures in the Food / Non-Food Retail Trade

To achieve this, it is necessary to consistently substitute (individual) shopping trips for purchasing products with high online affinity through efficiently organized delivery systems (cf., e.g., Daduna 2002) that will also be accepted by customers in terms of the routines and costs involved (cf., e.g., Punakivi / Saranen 2001). The focus here will be direct delivery (or, under certain circumstances, a delivery box system, if feasible), as a pick-up system runs counter to many essential customer benefits (e.g. relief from physical processing). Pick-up point delivery runs with minor acceptance in some places. However, these systems are still in a phase of testing (cf., e.g., as in the subproject within the M21 Project at DaimlerChrysler in Sindelfingen, Germany). The pick-up points are implemented at the workplace, and thus can be used on the way home from work.

Because products ordered online are supplied via local distribution warehouses and large-area shopping outlets, which simultaneously remain part of traditional retail structures, a concentration process can be anticipated in this segment with an increasing share of online shopping and correspondingly decreasing volume of turnover in traditional retail trade. To assure the (frequently short-term) coverage of requirements with products not seen as having online affinity, structures must be found with a strong spatial proximity to customers through *close-quarters shopping outlets*. A potential model in this regard is to construct well-positioned *neighborhood stores*, designed in the meaning of *convenience shops* (cf., e.g., Burgdorff 2000, p. 46ff; Daduna 2003). In this way, it can (and ultimately must) be assured that no additional traffic volume in motorized individual traffic arises based on any assortment-related division.

5.2 Changes in traffic volumes on the demand side

Many recent studies have sharply scaled back the high expectations for the traffic-related effects of online shopping that arose until the mid-1990s. Based on the assumption that the route to the retail store is eliminated for goods purchased electronically, Luley et al. (2002) calculate the effects of e-commerce on shopping traffic. Their model assumes that all goods that are purchased electronically will no longer need any physical shopping or transport activity on behalf of the consumer. The empirical input to assess the model stems from a survey in the region of Stuttgart, in southwest Germany. The survey asked – product by product - for individual physical shopping behaviour (means of transport used for shopping, place of discrete shopping activity and frequency of shopping), for e-commerce affinity (e-commerce experience and intentions) and for a self-estimation concerning online shopping behaviour by the year 2010. The results show that the traffic effects lag far behind the expectations. As a whole, the scope of kilometers driven for shopping purposes is reduced by only about 3%, while shopping trips are reduced by 2%.

Basic substitution effects can be presumed, though at a very low level, from other empirical studies. These studies were in part carried out with much smaller samples, and singled out specific e-commerce services, such as online banking (cf., e.g., Vogt / Lenz 2002, Zoche et al. 2002, pp 97). The substitution thesis is also supported through studies like that of Cairns (2003), which reach a quantification of potential effects by building models of the supply traffic and the customer traffic replaced thereby and comparing the two types of traffic expenses. It is likewise assumed in this method that the buying of goods over the Internet could permanently reduce physical purchase transactions by customers.

In contrast to these results, Casas et al. (2001) presume that electronic purchases do not entail any substitution effects. Based on data collected in 1999 during a traffic survey conducted among households in the region of Sacramento, California, the attempt was made to specify the typology of the "Internet shopper" in more detail. The study showed that persons who frequently shopped on the Internet also undertook numerous shopping trips.[2] Casas et al. (2001) concluded that no substitution took place, because the persons who use online shopping change their shopping behavior but not their mobility behavior, i.e. the number of shopping trips is not reduced. One possible explanation is offered by Gould / Golob (1997). They object to the assumption that the availability of product information is enough of a basis for shopping travel to be substituted through online shopping: "...because people enjoy some types of travel, and they shop for other than economic purposes, electronic home shopping could also generate additional travel and new types of in-store shopping activity" (Gould / Golob 1997). At any rate, the findings of Casas et al. (2001) support the thesis of the complementary or "add-on" effects arising from the use of online shopping.

As a whole, the studies on the relation between online shopping and traffic in recent years have become more comprehensive and differentiated. Various studies on consumer traffic behavior show that on the micro or individual level online shopping takes the place of the physical purchase for some, while representing an add-on for others or even inducing them to take additional trips (Bhat et al. 2001). It is in no way improbable that online shopping can fulfill these differing functions for one and the same person, depending on the product and the situations in which this marketing channel is used. The difficulty of determining the traffic effects based on changed shopping behavior is significantly compounded.

Expectations for some substitution of consumer traffic can be maintained anyway. In a recent study about the relation between transport behaviour and the use of information and communication technologies (ICT) 12,3% respectively 22,5% of the respondents say explicitly they use online shopping for time or travel saving (Nobis / Lenz 2003).

5.3 Effects of online shopping on customer and supplier traffic - An attempted netting of effects

Even invoking the currently available *empirical* findings on the effects of online shopping on traffic, statements on the overall effects primarily still remain estimates on the likelihood of developments. However, by including both customer and supplier traffic in these estimates, their reliability increases sharply. The assumptions made above regarding traffic effects can thus be specified more clearly.

---

[2] This statement by Casas et al. (2001) contradicts the findings of Luley (2003) to a certain degree. In a study based on empirical data about shopping behavior, Luley concludes that no clear relation exists between virtual mobility and individual physical mobility. Luley, however, follows a different methodological approach than Casaset al., in that his comparison is based on a typology of subjects corresponding to the diversity of shopping locations frequented by them. The frequency of purchases, in contrast, plays no role in his considerations.

A major assumption must be made: that the use of online shopping will increase in the future and will expand to a larger number of product segments. This assumption appears legitimate in light of the recent still unpublished results of an empirical survey conducted by the DLR Institute of Transport Research. Based on data collected from the city of Berlin, this survey provides evidence that the number of persons shopping online is increasing. In the case of Berlin today, already 31% of the population have experience with online shopping, which is slightly above the national average (cf., gfk webgauge, August 2001; www.gfk.de). At the same time, current and future online shoppers assume that the product range of their online purchases will increase (cf. Figure 5). Hence, substitution effects can indeed be expected, at least tendentially, though it can be presumed that these will hover at a low level. The success of substitution will depend on the extent to which online shopping replaces traffic-intensive purchases or saves trips related to their preparation. At the same time, substitution is only realizable if supply transports can be bundled, either within the existing structures and / or within the framework of newly developed routines. The success of this bundling will, in turn, depend on reaching a critical mass of shipments.

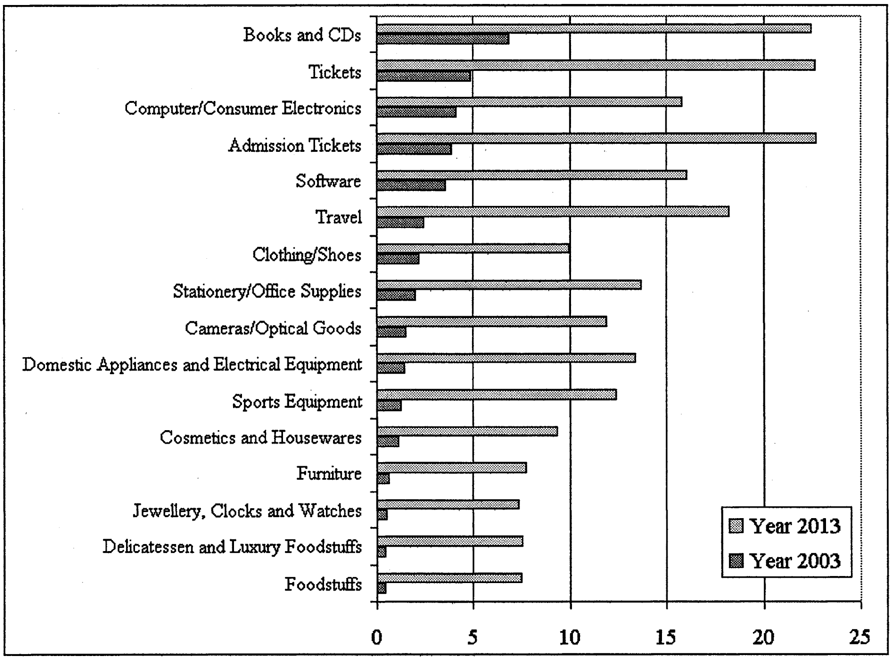


Fig. 5. Shares of purchases via online shopping in Berlin in 2003 and 2013 (source: DLR Institute of Transport Research, Berlin)

To the extent that existing structures of parcel services can be used, this critical mass is given, in the case of (industrial) direct marketing and the mail-order selling. Problematic, in contrast, is the food / non-food retail trade, where autonomous distribution systems must be organized due to the specific requirements (cf., e.g., Daduna 2002). Because customer orders are bundled here, a substitution effect can be expected provided the delivery location is located close enough to the supplier territories. The reason for such an assumption is that a large number of (individual) *shuttle trips* (between customer locations and shopping facilities) can be replaced through efficiently organized delivery tours.

6 Summary and Perspectives

Changes in traffic volumes on both the customer and supplier side are necessarily associated with the sustainable establishment of online shopping as a marketing channel. With the commercialization of the transport routines in the „last mile“, a shift of (logistical) activities from consumers to suppliers will necessarily take place. To what extent savings in mobility volumes will result cannot be quantified or can only be quantified conditionally. Two fundamental issues are of key importance at this point: the question of the length of transport routes and the efficiency of producing services in the logistical distribution area. Apart from the food / non-food retail trade, distances allowing an increase in mobility to be expected could arise through the elimination of locational ties in the stationary retail trade and through the internationalization of purchase decision-making processes. To what extent this will, in fact, occur depends on the evolution of service in distribution logistics. If it is possible to develop efficient structures in this area with bundling effects sufficient enough to serve the „last mile“ with sometimes atomistically distributed demand, the effects described in the substitution thesis could be realized.

The basic structures of many of the logistical service providers in the parcel and general cargo segment display a high degree of efficiency based on corresponding bundling potential and on efficient routines. Yet, the problems sometimes existing in the „last mile“ should not be overlooked in delivery routines. A key (and crucial) item in this context is the increasing lack of homogeneity in customer structures due to the growth in the share of non-commercial customers. Various attempts to arrive at acceptable and economically solid solutions with new, problem-specific strategies have failed, as the example of *sameday.com* has shown. This means that suitable procedures will have to be developed in this area in order to assure the necessary efficiency of routines. Moreover, the existing structures can only be transferred unrestrictedly to direct marketing and the mail-order trade, not to the food / non-food retail trade, which has a specific requirements profile in key areas. Completely new approaches must in part be developed for delivery routines (cf., e.g., Daduna 2002) that can satisfy the requirements of this segment. Delivery box systems and the integration of pick-up points (s. Pflaum 2003) will not take on any significant importance in this context, because the basis for economically viable realization is not given and the actual attraction of home delivery, comprehensive relief to consumers from the sometimes (physically) expensive activities of transporting goods, is missing.

In the food / non-food retail trade, negative effects could moreover arise with respect to the scope of mobility, as has been assumed in the complementarity thesis. In order to prevent these, the aforementioned structural changes in the retail trade are necessary. These can naturally not be realized in the short term, because the existing structures entailing stationary shopping outlets can only be adjusted in the medium or long term due to past investments and / or long-term lease agreements. Transitional phases will necessarily arise here, which may generate (temporary) additional mobility under certain circumstances.

In accordance with the induction thesis, a reduction of shopping expense will lead to an increase of available time, through which additional (politically undesired) mobility might be generated in the leisure area. But this should not be viewed as a necessary consequence, given that it cannot be said with certainty that these temporal gains will in fact be invested in activities necessitating additional expenditure in motorized individual traffic. Here, too, the empirical research is still only in its infancy.

Despite the widespread impression that the changes in mobility volumes induced through online shopping are largely determined by consumer behavior, we can find no justification for this statement either in the experience to date or from a theoretical point of view. Consumer mobility is doubtlessly *one* key factor, but should not be considered in isolation. Mobility will also be influenced by the anticipated changes in the locational structures of the retail trade (see above), because these structures can have effects on the choice of transport means (e.g. through the reduction of motorized individual traffic in some cases). The structure of delivery transport represents a further decisive factor (see above), as the potential for greater efficiency is evident when larger shipment volumes can be assumed. Decisive in this context is the question of the extent to which *cooperative models* can be realized when structuring delivery routines in order to achieve bundling effects at this level. Because logistic costs will be of increasing significance in the B2B-segment, the prospect indeed exists that changes will be achievable here (viewed over the long term), in some cases in connection with solutions for city logistics.

References

Bhat, C. / Sivakumar, A. / Axhausen, K. W. (2001): An analysis of the impact of information and communication technologies on non-maintenance shopping activities. Paper submitted to the 81st Annual Meeting of the Transportation Research Board Washington, DC, January 12-16, 2001; TRB ID Number 03-3290

Bretzke, W.-R. (1999): Smart Shopping im Internet - Industrie und Handel im Zeitalter des Electronic Commerce. in: Kopfer, H. / Bierwirth, C. (Hrsg.): Logistik Management - Intelligente I+K-Technologien. (Springer) Berlin et al., 221 - 243

Bretzke, W.-R. (2000): Electronic Commerce als Herausforderung an die Logistik. in: *logistik management* 2, 8 - 15

Bucklin, L. P. (1963): Retail strategy and the classification of consumer good. in: *Journal of Marketing* 27(1), 50 - 55

Burgdorff, F. / Getzmann, S. / Heinze, M. / Steinmüller, K. (2000): Online-Shopping und die Stadt. WerkstattBericht 28 / Sekretariat für Zukunftsforschung, Gelsenkirchen

Cairns, S. (2003): Mehr Verkehr durch Zustelldienste des Lebensmitteleinzelhandels? in: Jessen, J. / Lenz, B. / Roos, H. / Vogt, W. (Hrsg.): B2C Elektronischer Handel - Eine Inventur. (Leske & Budrich) Opladen, 265-294.

Casas, J. / Zmud, J. / Bricka, S. (2001): Impact of shopping via internet on travel for shopping purposes. Paper submitted to the 80th Annual Meeting of the Transportation Research Board Washington, DC, January 7-11, 2001; TRB ID Number 01-3393.

Dach, C. (2000): Die zukünftige Bedeutung des Business-to-Consumer E-Commerce - Ein Prognosemodell. in: Müller-Hagedorn, L. (Hrsg.): Zukunftsperspektiven des E-Commerce im Handel. (Deutscher Fachverlag) Frankfurt a.M., 175 - 225

Daduna, J. R. (2002): Distribution planning with specific delivery time restrictions for the handling of electronic customer orders in food / non-food retail trade. in: Klose, A. / Speranza, M. G. / Van Wassenhove, L. N. (eds.): Quantitative approaches to distribution logistics and supply chain management. (Springer) Berlin et al., 149 - 162

Daduna, J. R. (2003): Distributionsstrukturen und Verkehrsabläufe im Handel unter dem Einfluss des Online-Shoppings. in: *logistik management* 5, 12 - 24

de Koster, R. B. M. (2002): The logistics behind the enter click. in: Klose, A. / Speranza, M. G. / Van Wassenhove, L. N. (eds.): Quantitative approaches to distribution logistics and supply chain management. (Springer) Berlin et al., 131 – 148

Deutsche Verkehrswissenschaftliche Gesellschaft (DVWG) (2002): Verkehrliche Wirkungen von e-Business. Bergisch Gladbach.

Fichtner, C. / Engelsleben, T. (2001): Alternative Zustellkonzepte - Pakete von der Tankstelle. in: *Logistik heute* 23(9), 42 - 45

Frank, J. / Frechen, J. (1998): Freizeit und Einzelhandel - Ein Rezept für die Zukunft? - Voraussetzungen und Chancen. in: *Immobilien Zeitung* Nr. 20, 10 u. Nr. 21, 9

Gould, J. / Golob, T. F. (1997): Shopping without travel or travel without shopping? An investigation of electronic home shopping. in: *Transport Reviews* 17 (4), 355 - 376

Grem, F. (2001): Der Turmbau zu Dortmund. in: *dhf - deutsche hebe- und fördertechnik* 47(5), 10 - 12

Halder, G. / Habermann, U. / Schenk, W. / Valley, H. (1999): Zentrenkonzepte und Warentypologien - Komplementarität und Divergenzen. in: Priebs, A. (Hrsg.): Zentrale Orte - Einzelhandelsstandorte und neue Zentrenkonzepte in Verdichtungsräumen. Kieler Arbeitspapiere zur Landeskunde und Raumordnung 39, 23 - 36

Hirn, W. (2002): Vorbild Automobilindustrie - Wie das Internet die Beziehungen zwischen Herstellern, Kunden und Zulieferern effizienter gestaltet. in: *Netmanager* 2(1), 14 – 21

Holton, R. T. (1958): The distinction between convenience goods, shopping goods and speciality goods. in: *Journal of Marketing* 23(7), 53 - 56

Infas GmbH (2001): e-Business und Verkehrsaufkommen - Der Endkunde als Einflußfaktor. Ergebnisse einer Repräsentativerhebung im Auftrag des Deutschen Verkehrsforums im September / Oktober 2001, Bonn

Janz, O. (2001): Mehr Verkehr durch E-Commerce? - Eine Analyse der Auswirkungen des E-Commerce-Wachstums (B2C) auf den Verkehr. in: *Zeitschrift für Verkehrswissenschaft* 72, 48 - 69

Karg, M. (1998): Dell Computer Corporation - Wachstum durch Direktvertrieb. in: Tomczak, T. / Reinecke, S. (Hrsg.): Best Practice in Marketing – Erfolgsbeispiele zu den vier Kernaufgaben. (Thexis / Ueberreuter) St. Gallen / Wien, 49 – 64

Klaus, P. / König, S. / Pilz, K. / Voigt, U. (2002): Auswirkungen des elektronischen Handels (E-Commerce) auf Logistik und Verkehrsleistungen – Verknüpfung betriebswirtschaftlicher und volkswirtschaftlicher Informationen. Gutachten im Auftrag des Deutschen Bundestags. Berlin / Nürnberg

Kuchenbecker, M. / Siebel, L. (2001): Bleibt der Kühlschrank trotz E-Commerce leer? in: *Logistik heute* 23(10), 80 - 83

Lenz, B. (2003a): Verändert E-Commerce den [Einkaufs]verkehr? in: Ducar, D. / Rauh, J. (Hrsg.): Geographische Handelsforschung, Themenheft "E-Commerce", 87 - 99

Lenz, B. (2003b): Wer kauft schon im Internet? - Das Einkaufsverhalten von Online-Kunden und seine räumliche Relevanz. in: Jessen, J. / Lenz, B. / Roos, H. / Vogt, W. (Hrsg.): B2C Elektronischer Handel - Eine Inventur. (Leske & Budrich) Opladen, 64 - 80

Luley, T. (2003): Physische und virtuelle Mobilität - Modellbildung und empirische Befunde aus der Region Stuttgart. Stuttgarter Geographische Studien (in print)

Luley, T. / Bitzer, W. / Lenz, B. (2002): Verkehrssubstitution durch Electronic Commerce? - Ein Wirkungsmodell für die Region Stuttgart. in: *Zeitschrift für Verkehrswissenschaft* 73, 133 - 155

Merkel, H. / Kromer, S. (2001): Letzte Meile - Alte Idee im neuen Gewand. in: *Logistik heute* 23(11), 44 - 45

Merz, M. (2002): E-Commerce und E-Business - Marktmodelle, Anwendungen und Technologien. 2., aktual. u. erw. Aufl. (dpunkt.verlag) Heidelberg

Mindali, O. / Salomon, I. (2003): E-Retail and travel behavior - A conceptual model. in: http://www.stellaproject.org/FocusGroup1/Brussels2003/abstracts. doc, Zugriff am 26.5.2003

Müller-Hagedorn, L. (1993): Handelsmarketing. 2., überarb. u. erw. Aufl. (Kohlhammer) Stuttgart et al.

Nobis, C. / Lenz, B. (2004): Changes in transport behaviour by the fragmentation of activities. Paper presented at TRB Washington, 2004.

Palombo, P. / Theobald, A. (2000): Electronic Shopping - Das Versandhaus Quelle auf dem Weg in das Zeitalter der elektronischen Medien. in: Bliemel, F. / Fassod, G. / Theobald, A. (Hrsg.): Electronic Commerce. 3., überarb. u. erw. Aufl. (Gabler) Wiesbaden, 389 - 400

Pflaum, A. (2003): Die Zukunft des "E-Fulfilment" für Lebensmittel - Versuch einer Prognose. in: logistik management 5, 25 - 39

Punakivi, M. / Saranen, J. (2001): Identifying the success factors in e-grocery home delivery. in: *International Journal of Retail & Distribution Management* 29, 156 - 163

Reichwald, R. / Piller, F. T. (2000): Mass Customization-Konzepte im Electronic Business. in: Weiber, R. (Hrsg.): Handbuch Electronic Business. (Gabler) Wiesbaden, 359 - 382

Schögel, M. / Birkhofer, B. / Tomczak, T. (1999): Einsatzmöglichkeiten des Electronic Commerce in der Distribution. in: Tomczak, T. / Belz, C. / Schögel, M. / Birkhofer, B. (Hrsg.): Alternative Vertriebswege. (Thexis) St.Gallen, 288 - 308

Tietz, B. (1993): Der Handelsbetrieb. 2., neubearb. Aufl. (Vahlen) München

Vogt, W. / Lenz, M. (2002): Physischer Weg oder virtueller Kontakt? Einige empirische Befunde mobilitätsbezogener Wirkungen aus dem Bereich des E-Commerce. in: Forschungsgesellschaft für Straßen- und Verkehrswesen (Hrsg.): Heureka ´02. Optimierung in Verkehr und Transport. Tagungsbericht, 481-494

Wagener, R. / Jahn, D. / Lasch, R. / Lemke, A. (2002): Der Transportmarkt im Wandel - Endbericht zur Delphi-Studie. Bericht erstellt im Auftrag des Deutschen Verkehrsforums e.V., Berlin

Weber, J. / Schmitt, A. / Engelbrecht, C. / Knobloch, U. / Wallenburg, C. M. (2002): E-Commerce in der Logistik - Quantensprung oder Business as usual? (Haupt) Bern et al.

Wirtz, B. W. (2001): Electronic Business. 2,.vollst. überarb. u. erw. Aufl. (Gabler) Wiesbaden

Zoche, P. / Kimpeler, S. / Joepgen, M. (2002): Virtuelle Mobilität: Ein Phänomen mit physischen Konsequenzen? (Springer) Berlin et al.

Chapter 2

Vehicle Routing

A Real-time Vehicle Routing Model for a Courier Service Problem

Enrico Angelelli[1], Renata Mansini[2] and M. Grazia Speranza[1]

[1] Department of Quantitative Methods, University of Brescia, Italy
[2] Department of Electronics for Automation, University of Brescia, Italy

Abstract. The information needed to solve a vehicle routing problem may be not completely known ahead of time. Rather it may be dynamically revealed as time goes on. We consider a dynamic vehicle routing problem faced by a courier company where customer requests with service time windows arrive and have to be serviced on real time by a fleet of vehicles in movement. Differently from other dynamic routing problems motivated by the same courier service, we consider both pick-up and delivery requests and assume that customer requests cannot be refused but can be postponed to future shifts. A heuristic algorithm based on local search is proposed for the problem together with an illustrative example. Experimental analysis is in progress.

Keywords: Real-time vehicle routing problems, courier service, time windows.

1 Introduction

In this paper we consider the case of a courier company facing orders which come up continuously over time. The courier service is organized by geographical areas. Each area has a central hub which provides the vehicles for servicing the orders within the same area.

Each order settled to a hub consists of a pick-up and a delivery request (the parcel is collected from an origin and delivered to a destination). If an order consists of a delivery request that has a destination hub different from the hub of its pick-up request, then its transshipment is made overnight.

For a courier service, the time of the pick-up and delivery requests usually depends on the arrival time of the order at the call-center of the hub. The customer can freely set the deadline for the delivery of his/her parcel. However, the delivery within the morning of the day following the settlement of the order is satisfied only if the order is received by the call-center of the hub before a given time t. In such a case the pick-up is guaranteed within the end of the day in which the order was settled. If the order is released later than time t, then the delivery is no more guaranteed by tomorrow, but may be shifted till the day after tomorrow within the chosen deadline. This is due to the fact that the courier company is not able to guarantee the collection of

B. Fleischmann et al. (eds.), *Distribution Logistics*

the parcel within the day and such parcel may need a night to be transhipped to its destination hub, if different from the original one.

The service is organized in two non-overlapping shifts (morning and afternoon). Each request has a deadline established by the customer independently from the shift. However, if the pick-up request of an order takes place in a shift, the corresponding delivery cannot be accomplished before the successive shift when directed to a different hub. For this reason each hub simultaneously faces both types of requests, the pick-up and the delivery ones without coupling constraints. Moreover, all those requests whose service deadline is successive to the ending time of the current shift have not to be obligatorily serviced in the current shift and may be postponed to some later shift.

Some requests (usually the pick-up part of an order) may arrive on-line over time, while the delivery requests are typically known at the beginning of a shift. More precisely, each request is characterized by a release time. For a pick-up request which occurs in real time in a shift, the release time corresponds to the time at which the order is received by the call-center, while for a delivery request such time can be associated to the instant at which the corresponding pick-up is executed. Notice that all the requests, with release time in some earlier shifts, are known at the beginning of the current one. These are delivery requests of orders whose pick-up has taken place in an earlier shift or pick-up requests of orders received by the call-center in some earlier shift and not yet serviced.

The problem can be classified as a dynamic vehicle routing problem due to the uncertainty which comes from the occurrence of new requests which have to be assigned to vehicles in real time. The main objective of the problem is the optimization of the operational costs in the long period. This is obtained in the short period, namely in the shift, through the determination of a set of routes which maximizes the total value of the requests serviced by the vehicles within the shift duration.

Very recent works dealing with real time vehicle dispatching have appeared in the literature. In (Gendreau et al. 1999) the authors deal with a dynamic vehicle routing problem also motivated by a courier service application. For this problem the authors propose a tabu search algorithm implemented on a parallel platform. With respect to our problem the authors do not distinguish between pick-up and delivery requests. Moreover, in their problem a new request is rejected if it has no feasible insertion. In our case new requests might be postponed to future shifts but cannot be refused.

With respect to (Gendreau et al. 1999), in (Ichoua et al. 2000) the authors include diversion issues: once a driver is on a route to his next destination he can diverse his direction to serve new customer requests. Through empirical evaluation the authors show how to divert a vehicle from its current destination may result in potential savings. Our problem formulation also includes diversion issues.

Traditionally, dynamic problems have been handled by heuristic solution algorithms. One of the most common approaches has been the adaptation of well-known static algorithms to the problem (see Powell 1988, Gendreau and Potvin 1998, Ghiani et al. 2003), while some studies have introduced stochastic methods (see Bertsimas and Van Ryzin 1991, 1993). As far as successful applications of static algorithms to real dynamic problem are concerned, in (Savelsbergh and Sol 1998), the authors present a planning module (DRIVE) to be incorporated in a decision support system for a real transportation problem characterized by dynamic pick-up and delivery requests and vehicles with different capacities. The core of DRIVE is a Branch and Price algorithm for the general Pick-up and Delivery Problem (PDP). Finally, the readers interested to a general survey and introduction on dynamic problems are referred to (Psaraftis 1995).

In this paper, we propose a heuristic procedure to be used for handling both the a priori known and the on-line requests of a courier service.

The paper is organized as follows. In Section 2 we provide the complete problem description. Section 3 is devoted to the solution algorithm. Finally, in Section 4 an application of the proposed algorithm to a simple instance is shown as an illustrative example.

2 Problem Description

We will focus our attention on the problem handled by a single hub over time.

The courier service in a hub is organized in shifts which do not overlap each other. Let a_s and b_s denote the starting and ending time of the shift s (i.e. at a_s the vehicles leave the hub and at b_s they have to be back), while $b_s - a_s = h_s$ is the length of the shift. For each shift, the planner has to face two types of requests, the pick-up and the delivery. Each request r is characterized by a release time t_r and a hard time window $[e_r, l_r]$, where e_r is the earliest time for the service and l_r is the latest one.

A request time window can be either completely contained in a shift or may overlap consecutive shifts. If $l_r > b_s$ then the request r is *postponable* with respect to shift s, otherwise the request is *unpostponable.*

We will identify as *off-line requests* with respect to shift s all those requests whose order was settled in some earlier shift, i.e. if $t_r < a_s$. On the contrary, we define as *on-line request* with respect to shift s the request r such that $a_s \leq t_r < b_s$, that is the request arrives during the shift. Notice that both on-line and off-line requests (either deliveries or pick-ups) may be either postponable or unpostponable according to their time windows.

Delivery requests (either on-line or off-line) have to be assigned to a vehicle. The assignment cannot be modified during the current shift. This means that when the parcel is loaded on a vehicle the corresponding delivery must be accomplished by the same vehicle. It is worth noticing that on-line delivery requests can only arise when an order involves both the pick-up and the delivery within the same area. The on-line delivery request (postponable or unpostponable) is automatically assigned to the vehicle which will perform the pick-up.

The main objective of the problem is the optimization of the operational costs in the long period.

The time-dependent nature of the problem imposes to coordinate the decisions taken over all the shifts. Moreover, since the requests arrive on real-time, the decisions taken at different time instants have to be coordinated also during each shift.

Thus, the problem can be seen as structured on three stages:

- In the long term (over all the shifts) we have to define the service policy with the objective of minimizing the operational costs. At this aim a priority is assigned to every request at the beginning of each shift. The priority is a value depending on the request urgency. Such value may be modified over time and should increase as the request deadline is approaching.
- During the shift we have to choose the set of requests to be served and the requests to be put off to a successive shift as well as the route for each vehicle. Note that at this stage we cannot formulate a classical optimization problem since each decision, at least in principle, should be revised when a new on-line request is released. At this stage the minimization of the operational costs might result in a suboptimal policy implying that only the requests which cannot be postponed are served even when vehicles and time are available.
- As several on-line requests may be released in a short time, it is sensible to periodically revise the decisions (say every 10 minutes). At a given time (say 10:20 am) we record the on-line requests occurred in the previous period (between 10:10 and 10:20) and then we take the decisions on the basis of all the available information. This means that at this stage a classical optimization problem can be formulated and solved. A critical issue is the identification of the objective function which guides the long term minimization of the operational costs. We have chosen to maximize the total value (the sum of the priorities defined at the first stage) of the requests which can be feasibly served from the current time to the end of the shift.

From now on, we will focus on the latter stage. Let $M = \{v_1, ..., v_m\}$ be the set of vehicles available for the service. Let $P = \{p_0, p_1, ..., p_m\}$ be the position of the hub and of vehicles $v_1, ..., v_m$, in any time instant. We define as R_U and R_P the set of unpostponable and postponable requests, while $R = R_U \cup R_P$ is the total set of requests. Each request r has a positive value w_r. To each vehicle v_h is assigned a set D_h of delivery requests, where $\cup_{h=1,...,m} D_h \subseteq R$ is the total set of the delivery requests.

We define the problem on a graph $G = (V, A)$, where $V = P \cup R$ is the set of vertices and A is the set of arcs.

Each arc $(i, j) \in A$ is weighted with nonnegative values d_{ij} and t_{ij} which indicate the distance and the travel time from vertex i to vertex j, respectively. For simplicity it is assumed that the travel time t_{ij} includes the service time at vertex i.

The objective is to find a set of m routes which maximizes the total value of the serviced requests, while satisfying the following constraints:

- the route of vehicle v_h must originate in p_h and terminate in p_0 (the hub) for each $h = 1, ..., m$;
- each request can be serviced at most once by only one vehicle;
- each delivery request in D_h can be served only by vehicle v_h;
- each unpostponable request $r \in R_U$ has to be satisfied within its time window $[e_r, l_r]$. The vehicle can arrive before e_r and in such case it has to wait but cannot serve the request later than l_r;
- each postponable request $r \in R_P$ can be satisfied only within its time window $[e_r, l_r]$. The vehicle can arrive before e_r and in such case it has to wait but cannot serve the request later than l_r;
- each route has a duration constraint equal to the length of the shift.

Note that, in the problem formulation, we do not directly consider operational costs minimization. However, the chosen maximization objective together with the constraint on routes duration leads also to the optimization of the total travel time.

In a dynamic framework we have to face the occurrence of dynamic events, i.e. the arrival of new requests. During the time frame between the arrival of two consecutive requests, the vehicles might have changed their position, while some requests may have been served. This is why, when dynamic events occur, the data of the problem have to be updated and the routes re-optimized.

In the following section we explain the solution algorithms introduced to tackle the described dynamic problem.

3 A Solution Algorithm

The present section is devoted to the description of the main algorithm and its subroutines.

The handling of the routes is based on the insertion (extraction) of single requests in (from) one route. Such operations are implemented as a personal version of the GENIUS algorithm proposed for the TSPTW by Gendreau et al. (1998). The algorithm consists of a tour construction phase (GENI) followed by a re-optimization phase (US) and is constructed so to always preserve time windows feasibility. The GENI algorithm is based on a set of rules to insert, one at the time, a set of vertices into a tour which starts from and ends to the depot. The insertion is performed after the evaluation of a number of different insertion rules. Every time the algorithm is unable to further insert any of the remaining vertices, it extracts a vertex from the tour, appends it to the list of vertices still to be inserted and starts again to insert. The extraction of a vertex is performed by applying the best among different extraction rules.

Our implementation of GENI solves a variant of the classical TSPTW looking for a path instead of a tour. Indeed, it builds a path from an initial position p_h (current position of vehicle v_h) to a final destination p_0 (the hub). Moreover, while the original implementation finds either a solution visiting all the vertices or fails, our implementation always returns a path visiting a set of serviced vertices and provides a list (possibly empty) of excluded vertices. We implement the insertion/extraction rules in two routines identified as *Insert* and *Extract*, respectively. These routines have two arguments: a set of routes and a set of vertices (requests). If a single vertex and a single route are passed to the *Insert* or *Extract* routines, the GENI algorithm is applied.

If a single vertex has to be inserted into (extracted from) a set of routes, then its insertion (extraction) cost is first evaluated for each route and then the vertex is inserted into (extracted from) the route with minimum insertion (extraction) cost. The insertion (extraction) cost is defined as the difference between the length of the route after the insertion (extraction) and the length before. The insertion cost is expected to be positive for insertions and negative for extractions.

```
Insert(S, r)
    INPUT: a set of routes S and a request r.
    OUTPUT: success or failure and the set of augmented routes S.
    s* = null
    δ* = 0
    S' = copy of S
    for each s' ∈ S'
        c = length(s')
        apply GENI to insert r in s'
        if insertion is successful
            δ = length(s') − c
            if (s* == null or δ < δ*), {δ* = δ; s* = s' }
    end for
    if ( s* == null), return failure
    else
        update S
        return success
end Insert
```

Insert(S, I)

INPUT: a set of routes S and a list of requests I.
OUTPUT: the set of augmented routes S and the list I
of requests failed to be inserted.
for $i = 1$ to $|I|$
$r =$ first request in I
remove r from I
if $(Insert(S, r) == failure)$,append r to I
end for
end Insert

When the insertion (extraction) of many vertices is required, the order in which the insertions (extractions) are performed can drastically affect the final result. Thus, we always consider ordered lists of requests. In particular, we have considered two sorting criteria: *sorting by value* and *sorting by cost.* In the first case, the value of the requests is the primary key for the sorting, while the insertion cost is used to order requests with equal values. In the second type of sorting, the role of the two keys is reversed. Notice that the two keys are independent and can produce quite different results. The values will be sorted in non increasing order for insertion (requests with big value are inserted first) and non decreasing order for extraction (requests with small value are extracted first). The insertion or extraction costs will always be sorted in non decreasing order. The proper sorting criterion will be used according to the situation. For example, when we aim at increasing the total value of the requests in the solution, we order the requests by value. On the other hand, when we aim at inserting a set of unpostponable requests in order to obtain a feasible solution, we order them by cost.

At the second stage, i.e. during each shift, an algorithm, *ShiftRoutine*, is run over time to provide the routes construction in the shift. As a first step such routine provides an initial solution $\underline{x}$ based on the lists of the postponable and unpostponable (namely, R_P and R_U) requests available at the beginning of the shift. Then, the routine consists of a main loop where the two lists, R_P and R_U, are updated with the new requests. The current solution $\underline{x}$ is made feasible (all requests in R_U are inserted into the current solution) and then it is improved. A fixed amount of time ($OptTime$) is dedicated to the execution of the loop which is repeated until the end of the shift. If the execution of the loop ends early, then the routine stops and waits until the end of the fixed time frame $OptTime$. The function $WallClock()$ gives the current time.

ShiftRoutine

$\underline{x} = InitialSolution()$
while $(WallClock() + OptTime \leq b_s)$
$InitialTime = WallClock();$
update sets R_P and R_U

update locations P at time $InitialTime + OptTime$
// re-optimize the routes
$\underline{x}' := MakeFeasible(\underline{x})$
$\underline{x} := VNLocalSearch(\underline{x}')$
wait until $(WallClock() \geq InitialTime + OptTime)$
end while
end ShiftRoutine

Notice that when $WallClock() + OptTime > b_s$ the loop is not iterated and the vehicles follow the current solution to the end. However, *OptTime* is meant to be "small" and the excluded requests are those with release time close to the end of the shift. In practice it is sensible to assume that such requests will be postponable for the current shift. Moreover, at each iteration a guess on future position of the vehicles has to be made. Such guess is difficult in practice. However, using small values of *OptTime* should help in reducing estimation errors. Such errors will not propagate if, at each iteration of the *ShiftRoutine*, the actual position of the vehicles is considered.

Let us analyze the *ShiftRoutine* components in more detail. As already mentioned, at the beginning of a shift s a set of off-line requests are known. These requests have to be assigned to the vehicles before the leave from the depot. While off-line pick-up requests may be assigned to different vehicles during the shift, off-line delivery requests have to be serviced by the vehicle to which the parcels to be delivered are physically assigned by the initial solution.

The initial routes construction is provided by the following *InitialSolution* procedure. The geographical area is partitioned into m identical cones (we work in the Euclidean plane) with vertex in the hub. For each cone $h = 1, ..., m$ the initial position p_h of vehicle v_h is set as p_0, a path from p_h to p_0 is initialized and all the requests falling inside the cone are inserted by value: first the unpostponable requests and then the postponable ones.

The call to the *MakeFeasible* routine may be necessary in two cases. At the beginning of the shift, the initial solution could be infeasible because it does not include some unpostponable requests. The second case occurs when the current solution becomes infeasible due to the arrival of some unpostponable requests. The routine works in two phases. It first tries to insert the set of unpostponable requests, which are not yet serviced, in the current solution. If some requests cannot be inserted, the requests already in the routes are rearranged in order to minimize the total travelled length (routine *LocalSearch_Cost*). After a fixed number of attempts (*MaxFailure*), the routine goes on to the second phase. In this phase the postponable requests are extracted (routine *Extract*) one at a time in order to leave room for the insertion (routine *Insert*) of the unpostponable ones. If all postponable requests are extracted and some unpostponable ones are still out, the routine fails.

The call to the insert routine $Insert(\underline{x}, I)$ returns the solution $\underline{x}$ augmented by some of (possibly all) the requests in I which could be successfully inserted in the routes. The list I is returned with the requests which could not be inserted into the solution. If $I = \emptyset$ then all the requests have been feasibly inserted in the solution $\underline{x}$.

MakeFeasible($\underline{x}$)
  INPUT: a solution $\underline{x}$.
  OUTPUT: either a new feasible solution $\underline{x}'$ or *failure.*
  $failureCounter = 0$
  I = requests of R_U not in solution $\underline{x}$
  sort I by non decreasing insertion cost
  $MaxFailure = |I|/2 + 1$
  $\underline{x}' = \underline{x}$
  // Phase 1
  repeat
    $Insert(\underline{x}', I)$
    if ($I = \emptyset$), return $\underline{x}'$
    else // some requests could not be inserted
      $failureCounter++$
      if ($failureCounter \geq MaxFailure$), **exit loop**
      $LocalSearch\_Cost(\underline{x}')$
  end repeat
  // Phase 2
  L_P = list of all postponable requests in $\underline{x}'$
  L_U = list of all unpostponable pick-up requests in $\underline{x}'$
  sort L_P and L_U by non decreasing extraction cost
  append L_U to L_P and make a single list L
  while ($|L| > 0$)
    r = first request in L
    remove r from L
    $Extract(\underline{x}', r)$
    if (r is unpostponable), append r to I
    $Insert(\underline{x}', I)$
    if ($I = \emptyset$), return $\underline{x}'$
  end while
  return *failure*
end MakeFeasible

The core of the *ShiftRoutine* is the variable neighborhood heuristic *VNLocalSearch*. Typically, local search heuristics get stuck in a local optimum. In order to move from a local optimum, an alternative solution is selected in a neighborhood of the current one and the local search is started from scratch. If the alternative solution is too close to the current one, the local search is likely to get back to the same local optimum. On the other hand if the alternative solution is too far from the current one, the search could move away from a promising area. The aim of a variable neighborhood scheme is to select the alternative solution in neighborhoods with variable radius. If intensification is required the initial radius will be minimum and then will be increased as the local search fails. If diversification is required, the initial radius will be maximum and then decreased as the local search fails. The variable neighborhood search scheme was first proposed by Miladenovic and Hansen (1997) and applied in various cases (see for instance Hertz and Mittaz 2001).

The *VNLocalSearch* starts with the radius set to the minimum value. At each iteration of the *VNLocalSearch* procedure an alternative solution $\underline{x}'$ in the neighborhood of the current solution $\underline{x}$ is selected. Then a local search is performed starting from $\underline{x}'$. If the local search fails to find a solution better than $\underline{x}$, then the radius of the neighborhood is increased and a new alternative solution $\underline{x}'$ is selected. If the local search succeeds, the current solution is updated and the radius is set to its minimum value. The solution neighborhood $N_k(\underline{x})$ of radius k is defined as the set of solutions obtained by removing k postponable requests and inserting as much as possible of the requests which were not included. In our implementation we randomly select one neighbor by means of the *Neighbor* routine. Such routine sorts the list of postponable requests in the current solution according to the non decreasing order of the extraction costs; then k requests are randomly selected out of the first $2k$. The selected requests are then extracted and labelled as tabu. Finally, the *Insert* routine is executed with the list of postponable requests which are not tabu.

VNLocalSearch($\underline{x}$)
INPUT: a solution $\underline{x}$.
OUTPUT: a final solution $\underline{x}^*$.
k_{max} = maximum value of the radius
$k = 1$
$\underline{x}^* = \underline{x}$
while $(k \leq k_{max})$
  $\underline{x}' = Neighbor(\underline{x}^*, k)$
  $LocalSearch\_Value(x')$
  if $(z' > z^*)$ // z', z^* are the values of solutions $\underline{x}', \underline{x}^*$
    $\underline{x}^* = \underline{x}'$
    $k = 1$
  else
    $k := k + 1$

end if
end while
return $\underline{x}^*$
end VNLocalSearch

The *LocalSearch_Value* routine works on two levels. At the first level it tries to increase the value of the solution, while at the second level it tries to reduce the cost of the routes with the aim to provide additional time for adding more requests and, thus, increase the value of the solution.

LocalSearch_Value($\underline{x}$)
INPUT: one solution $\underline{x}$.
OUTPUT: a final solution $\underline{x}^*$.
$\underline{x}^* = \underline{x}$
z^* = value of $\underline{x}^*$
do
$z' = z^*$
$LocalSearch\_Cost(\underline{x}^*)$
sort R_U and R_P by decreasing values
$Insert(\underline{x}^*, R_U)$
$Insert(\underline{x}^*, R_P)$
z^* = value of $\underline{x}^*$
while $(z^* > z')$
end LocalSearch_Value

The routine *LocalSearch_Cost* starts with the re-optimization of each route: each request is, in turn, extracted and re-inserted in the same route. Then, the list of all the pick-up requests already serviced in the current solution is sorted by extracting cost. The first q requests are extracted and re-inserted in the solution by means of the routine *Insert*. The list of requests for the *Insert* routine is ordered according to the extraction order. Such procedure is repeated for values of q from 1 to 4. Finally, the *LocalSearch_Cost* routine moves some of the pick-up requests from a route to another route according to the following criterion. For each request i (either pick-up or delivery) in the current solution, the pick-up request j which minimizes the quantity

$$\widetilde{d}_{ij} = d_{ij} + v \cdot max(e_j - e_i - t_{ij}, 0),$$

is computed, where v is a parameter which estimates the average speed (a similar quantity was proposed in Wolfler Calvo 2000). The first five pairs (i, j) with smaller value of $\widetilde{d}_{ij}$ are selected and the j requests are extracted from the solution. Then each j request is inserted in the route containing its corresponding i request. If the new solution is feasible and has a total length less than the current one, then the current solution is updated and the procedure is iterated. Otherwise the routine stops.

4 An Illustrative Example

In the following we provide a simple instance to show how the algorithm works as an illustrative example.

We consider the case of a hub with only two vehicles, i.e. $m = 2$. Requests are distributed over the Euclidean plane where the hub is located at the origin. The distances are Euclidean. The requests have the same value and the time needed to travel a unitary distance is unitary. The shift is assumed to start at time 0 and end at time 44.

Table 1 shows the main information associated to each request: the first column identifies the request, while the second one shows its type (a pick-up (P) or a delivery (D) request). The third column provides the request coordinates in the plane, the fourth one shows its time window measured in units of time and, finally, the last column indicates the release date representing the point in time in which the request occurs. If the release date is null this means that the request was one of the off-line requests available at the beginning of the shift. Request 0 indicates the depot and its coordinates. The time window of the depot determines the duration of the shift. All the requests are unpostponable, but request 4 whose time window closes at time 88.

Table 1. Computational example: the data.

Request r	Pick-up/ Delivery	Coordinates	Time window	Release date t_r
0 (depot)		(0,0)	[0,44]	
1	D	(0,8)	[8,9]	0
2	P	(8,11)	[30,33]	0
3	D	(9,-5)	[19,21]	0
4	P	(0,-5)	[5,88]	0
5	P	(8,8)	[25,33]	1
6	P	(3,-3)	[5,6]	2
7	P	(4,-3)	[6,7]	2
8	P	(8,15)	[26,30]	3

In Table 2 the distance (travel time) matrix is shown.

Figure 1 provides the initial solution where all the requests known at the beginning of the shift are assigned to the two vehicles. Pick-up requests are represented as black dots while delivery requests as white ones. The vehicle which serves requests 1 and 2 has clearly to wait after the service of request

Table 2. Distance (travel time) matrix.

	0	1	2	3	4	5	6	7	8
0	0	8.00	13.60	10.30	5.00	11.31	4.24	5.00	17.00
1	8.00	0	8.54	15.81	13.00	8.00	11.40	11.70	10.63
2	13.60	8.54	0	16.03	17.89	3.00	14.87	14.56	4.00
3	10.30	15.81	16.03	0	9.00	13.04	6.32	5.39	20.02
4	5.00	13.00	17.89	9.00	0	15.26	3.61	4.47	21.54
5	11.31	8.00	3.00	13.04	15.26	0	12.08	11.70	7.00
6	4.24	11.40	14.87	6.32	3.61	12.08	0	1.00	18.68
7	5.00	11.70	14.56	5.39	4.47	11.70	1.00	0	18.44
8	17.00	10.63	4.00	20.02	21.54	7.00	18.68	18.44	0

1 and before the service of request 2. Similarly, for the vehicle which serves requests 3 and 4.

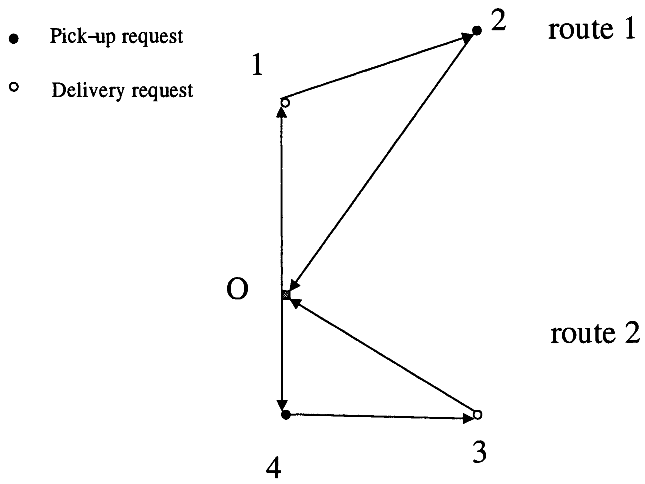


Fig. 1. Initial solution at time 0.

Figures from 2 to 4 show the on-line arrivals of the requests from 5 to 8 and the corresponding changes produced on the two routes at times 2, 3 and 4. The dotted lines in Figures 2-4 indicate the distances already travelled by the vehicles.

We assume that the interval time ($OptTime$) dedicated to the execution of the loop in $ShiftRoutine$ has been set to 1.

From time 0 to time 1 the algorithm tries to re-optimize the initial solution. At time 1 the request 5 arrives. Then, according to $ShiftRoutine$, such request is inserted into the list R_U and the position of the vehicles is

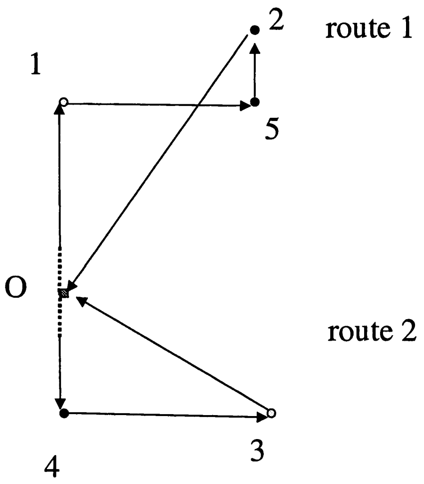


Fig. 2. New solution at time 2 after the release of the request 5 at time 1.

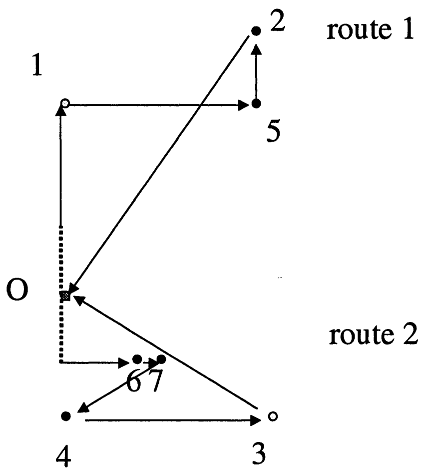


Fig. 3. New solution at time 3 after the release of the requests 6 and 7 at time 2.

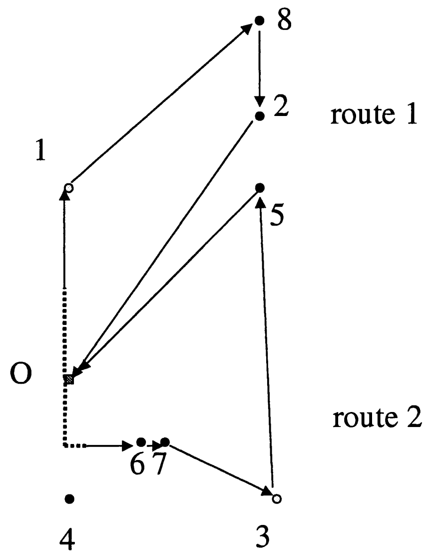


Fig. 4. New solution at time 4 after the release of the request 8 at time 3.

updated at time 2. This due to the fact that by the end of the re-optimization the vehicles will have been travelling for 1 unit of time. The result of this re-optimization is shown in Figure 2 where the request 5 is inserted in route 1 by the first phase of the *MakeFeasible* routine. Notice that the same request could have been inserted in route 2 without excluding the request 4, but at a larger cost. Similarly, in Figure 3, the new requests 6 and 7 which arrived at time 2 are added to route 2 at time 3.

Figure 4 shows that at time 4 the request 8 is inserted into route 1. Notice that request 8 cannot be inserted in any route by the first phase of *MakeFeasible*. The second phase of such routine first extracts request 4, but, still, request 8 neither can be inserted into route 1 nor into route 2. Thus, the unpostponable request 2 is extracted. Notice that request 5 presents a larger insertion cost, while request 3 cannot be extracted because it is a delivery. Finally, request 5 is extracted and only at this point request 8 can be inserted into route 1. Immediately after the insertion of request 8, requests 2 and 5 are successfully re-inserted into route 1 and route 2, respectively. Routine *MakeFeasible* ends with a feasible solution which does not include request 4. Request 4 is postponable and cannot be inserted in any route. It will be served in the next shift.

5 Conclusions

In this paper we have analyzed a real-time vehicle routing problem motivated by a courier service application where pick-up and delivery requests arrive on-line to a call center. We have structured the problem in three stages where the first two stages require decisions coordination over time, while at the third stage a static optimization problem is solved. While in the long term the objective is the minimization of the operational costs, in the short term, i.e. in the static optimization problem, a different objective has been defined as the maximization of the total value of the served requests. The static problem is solved by a variable neighborhood local search heuristic. During the experimental analysis, which is currently in progress, the values of the requests will be defined and the suitability of the chosen objective function for the static problem will be tested.

References

Bertsimas, D. / Van Ryzin, D. (1991): A stochastic and dynamic vehicle routing problem in the euclidean plane. *Operations Research*, 39:601–615.

Bertsimas, D. / Van Ryzin, D. (1993): Stochastic and dynamic vehicle routing in the euclidean plane with multiple capacitated vehicles. *Operations Research*, 41:60–76.

Gendreau, M. / Guertin, F. / Potvin, J. Y. / Taillard, E. (1999): Parallel tabu search for real-time vehicle routing and dispatching. *Transportation Science* 33:381–390.

Gendreau, M. / Hertz, A. / Laporte, G. / Stan, M. (1998): A generalized insertion heuristic for the traveling salesman problem with time windows. *Operations Research*, 43:330–335.

Gendreau, M. / Potvin, J.Y. (1998): Dynamic vehicle routing and dispatching. In: T. G. Crainic and G. Laporte (eds.), *Fleet Management and Logistics*, Kluwer Academic Publishers, pp. 115–126.

Ghiani, G. / Guerriero, F. / Laporte, G. / Musmanno, R. 2003): Real-time vehicle routing: solution concepts, algorithms and parallel computing strategies. *European Journal of Operational Research*, 151:1–11.

Hertz, A. / Mittaz, M. (2001): A variable neighborhood descent algorithm for the undirected capacitated arc routing problem. *Transportation Science*, 35:425–434.

Ichoua, S. / Gendreau, M. / Potvin, J. Y. (2000): Diversion issues in real -time vehicle dispatching. *Transportation Science*, 34:426–438.

Mladenovic, N. / Hansen P. (1997): Variable neighborhood search. *Computers and Operations Research*, 24:1097–1100.

Powell, W. (1988): A comparative review of alternative algorithms for the dynamic vehicle allocation problem. In: B. L. Golden and A.A. Assad (eds.), *Vehicle Routing: Methods and Studies*, North Holland, Amsterdam, pp. 249–291.

Psaraftis, H. M. (1995): Dynamic vehicle routing: status and prospects. *Annals of Operations Research*, 61:143–164.

Savelsbergh, M. W. P. / Sol, M. (1998): DRIVE: dynamic routing of independent vehicles. *Operations Research*, 66:474-490.

Wolfler Calvo R. (2000): A new heuristic for the travelling salesman problem with time windows. *Transportation Science*, 34:113-124.

Collection of Waste with Single Load Trucks: A Real Case

Claudia Archetti and Maria Grazia Speranza

Department of Quantitative Methods, University of Brescia, Italy

Abstract In this paper we consider a real case problem of waste collection in the county of Brescia, in the North of Italy, which we call the 1-Skip Collection Problem (1-SCP). This is a problem where a fleet of vehicles must collect a number of skips containing waste and transport them to one among different plants chosen on the basis of the kind of waste contained in the skip. Each vehicle has a capacity of one skip and it starts and ends its tour at the depot. Each time a vehicle collects a skip, it has to go to a plant and empty it. A number of constraints and problem characteristics have to be considered, as time windows for the customers and the plants, shift-time, different kinds of skips, number of drivers available to carry out the service and priorities assigned to the customers that have to be served. The objective is to minimize the total cost of the service given by the sum of the cost of the drivers engaged to carry out the service, the cost of the extra time and the penalty cost paid if a customer is not served. A heuristic algorithm to solve the real case problem is presented. The algorithm first constructs a feasible solution by means of the nearest neighbor algorithm and then improves it. The computational results show that the solution of the algorithm is much better than the solution applied by the firm which carries out the service since it serves a larger number of skips with a smaller number of drivers.

Keywords. Vehicle routing, heuristic algorithms, waste collection

1 Introduction

In this paper we analyze a real case problem of waste collection in the county of Brescia, in the North of Italy. We call this problem the 1-Skip Collection Problem (1-SCP) where a fleet of vehicles, which have all a maximum capacity of one skip, must collect a number of skips located at the customers. The skips contain waste of different kinds and each vehicle must transport an empty skip to a customer and collect a full one, go to a special plant to empty the skip and then continue with the service of a new customer (or the same customer if he has more than one skip to be collected). There are different plants that accept specific kinds of waste. A number of constraints and problem characteristics have to be considered, such as time windows for the customers and the plants, shift-time, different kinds of skips, number of drivers available to carry out the service and priorities assigned to the customers that have to be served. The objective is to minimize the total cost of the service given by the sum of the cost of the drivers engaged to carry out

B. Fleischmann et al. (eds.), *Distribution Logistics*

the service, the cost of the extra time and the penalty cost paid if a customer is not served.

A problem similar to the 1-SCP is analyzed in Ball et al. (2000) and is called the *Rollon-Rolloff Vehicle Routing Problem* (RRVRP). This problem has been introduced in the literature only few years ago and it belongs to the class of the *Vehicle Routing Problems* (for a survey see Ball et al. (1995), Cordeau et al. (2002) and Toth and Vigo (2002)). The RRVRP is a problem where there is a number of vehicles with a capacity of one skip that must move skips placed in different locations and transport them to a single plant. At the disposal plant, full skips are emptied and empty skips are loaded on the vehicles. A problem in which different plants are available is analyzed in De Meulemeester et al. (1997) where the authors model and solve a RRVRP with pick-up and delivery requirements with different kinds of customers and skips characterized by different volumes.

In the next section, we give a description of the 1-SCP. In Section 3 we describe the algorithm we have proposed to solve the 1-SCP. In Section 4 the data of the real case instance are presented together with the computational results and a comparison between the solution of our algorithm and the solution applied by the firm which carries out the service. Finally, in Section 5, some conclusions are presented.

2 The 1-Skip Collection Problem

In the 1-SCP a fleet of vehicles has to collect a number of full skips located at different customers and transport them to different plants. Each time a full skip is collected from a customer an empty one must replace it. A set of constraints is imposed:

1. **Number of drivers:** there is a bound on the number of drivers that can carry out the service. For each driver engaged a cost is paid.
2. **Capacity constraint:** each vehicle can transport only one skip at a time.
3. **Kind of waste:** there are different kinds of waste and each full skip has to be transported to a plant which accepts the kind of waste it contains. Each plant accepts only one kind of waste. There can be different plants for each kind of waste and a driver can normally choose one among all of them, except for the case of the customers that require that their skip is transported to a particular plant. A full skip contains only a single kind of waste and there is no change-over cost when the kind of waste in a skip is changed (e.g., cost of cleaning the skip).
4. **Shift-time:** the shift-time gives the maximum regular duration of a route. It is possible that a tour lasts more but in this case there is an extra cost for the time that exceeds the shift-time. Moreover, there is a constraint on the maximum extra time.

5. **Kind of skips:** there are different kinds of skips characterized by the volume and a vehicle must collect skips of the same kind during the entire tour. The volume of the skip must fit within the volume of the vehicle, i.e., a vehicle can transport only one kind of skip. Moreover, there can be customers which are the owners of their skip and require the same skip back after it has been emptied.
6. **Time windows:** each customer, and each plant, can be visited only if the arrival time of the vehicle is compatible with the time window of the customer, and the plant. If a customer, or a plant, is closed when the vehicle arrives, the vehicle can wait till the opening time.
7. **Priorities:** there are two groups of customers: the first one is formed by those customers that have a fixed demand, i.e., they have to be served at a constant frequency, while the second group is formed by those customers who require the service as soon as their skip (or skips) is full and thus at a variable frequency. The customers who have a fixed demand have a higher priority than the customers of the second group. The latter must be served on the basis of the arrival time of their request. There is a bound on the maximum number of days a customer can wait before being served which is the same for all the customers of the second group. Two different penalty costs are established when the service of a request is delayed and are proportional to the number of days a customer has to wait from the day of the request (or from the day corresponding to the fixed frequency) to the day of the service. The first cost c_{p_1} is paid when the service of a customer of the second group is delayed for a number of days lower than the bound. The second cost, c_{p_2}, is paid when the delay concerns a customer with a fixed demand or a customer that has already waited the maximum number of days. Both costs c_{p_1} and c_{p_2} are daily costs and thus are paid for each day of delay. We have $c_{p_2} > c_{p_1}$.
8. **Service time:** the operations of loading and unloading the skips require a certain time which has to be considered each time a vehicle collects a skip from a customer or empties a skip.

Thus, each vehicle starts its tour from the depot with an empty skip and goes to a customer which is open when the vehicle arrives, unloads the empty skip and loads a full skip of the same volume. Then, it reaches one of the plants which accepts the kind of waste contained in the skip, empties the skip and leaves with the empty skip. Afterwards it can serve a new customer, or the same customer if there are more skips to be served. The vehicle has to return to the depot when the shift-time is over. It is assumed that at the depot there is an unlimited number of empty skips of all volumes. There can be customers which are the owners of their skips: in this case, the vehicle arrives to the customer with an empty skip which is unloaded and a full skip is loaded and transported to the plant where it is emptied. Then the vehicle returns to the customer, it unloads the skip and it leaves with the empty skip it had when it arrived. If a driver returns to the depot without having emptied the last skip (because the plant was closed or because it has no more time to empty the skip), then the skip must be emptied in the next shift as soon as possible. There is a cost associated to each driver engaged to carry out the service.

The problem can be represented on a graph $G = (V, A)$ where the vertices represent the depot (vertex 0), the skips and the disposal plants. A customer with k skips generates k vertices in the graph. We denote by n the total number of skips, by m the number of disposal facilities and d the number of different kinds of waste contained in the skips. A vertex is associated to each skip and vertices $i = 1, ..., n$ represent the skips while vertices $i = n+1, ..., n+m$ represent the disposal plants. Each skip has the priority and the time window of the corresponding customer. Thus, $|V| = n + m + 1$. A parameter $k_i = 1, ..., d$ is associated to each vertex (except vertex 0) and indicates the number corresponding to the kind of waste contained in the skip (for $i = 1, ..., n$) or accepted by the plant (for $i = n+1, ..., n+m$). We denote by d_{ij} the distance from i to j and we associate a weight c_{ij} to each arc (i, j):

$$c_{ij} = \begin{cases} d_{ij} & \text{if } (i,j) \in S \\ \infty & \text{otherwise} \end{cases}$$

where

$S = \{(i, j) \mid (i = 0) \text{ or } (j = 0) \text{ or } (i = n+1, ..., n+m \text{ and } j = 1, ..., n) \text{ or } (i = 1, ..., n \text{ and } j = n+1, ..., n+m \text{ and } k_i = k_j) \text{ or } (i = 1, ..., n \text{ and } j = 1, ..., n \text{ and customer } i \text{ owns his skip})\}$.

The objective is to find the tours of the vehicles such that all the constraints are satisfied and the objective function is minimized. The objective function is given by the sum of three terms:

- the cost of the drivers engaged to carry out the service;
- the cost of the extra time;
- the penalty cost.

3 The Algorithm

We have designed a heuristic algorithm for the 1-SCP which first finds a feasible solution and then improves it. On the basis of the requests arrived during the day, the algorithm plans the service for the day after. The general structure of the algorithm is the following:

- *Phase 1*: choose a subset of skips with high priority;
- *Phase 2*: construct a feasible solution on the subset of skips determined in *phase 1*;
- *Phase 3*: improve the solution found.

In *phase 1* the algorithm selects a subset of skips with the highest priority. The number of selected skips is given by the average number of skips which can be served during the day. On this subset of skips the algorithm constructs a feasible solution. This forces the algorithm to first serve the skips with the highest priority, thus reducing the penalty costs.

The construction of a feasible solution (*phase 2*) is based on the nearest neighbor algorithm. The algorithm tries to serve all the requests determined in the previous phase. The problem is divided in subproblems on the basis of the volume of the skips since skips with different volumes must be served by different vehicles. For each subproblem, the routes are constructed sequentially: starting from the first vehicle, the algorithm chooses to serve the skip nearest to the depot. The choice is made only on the set of skips corresponding to customers which are open when the vehicle arrives. After having collected a skip, the vehicle goes to the nearest plant which is open when it arrives and empties the skip. It then goes to the nearest customer which is open and the procedure is repeated till the end of the shift-time. If the service of a skip requires to exceed the shift-time, the algorithm evaluates the possibilities of serving it during the extra time or in the next shift at the ordinary labor cost and it chooses the best solution. Afterwards, it starts constructing a new route and it repeats the previous procedure. Once a feasible solution is constructed, it is possible that some skips which have to be served during the day (that is, skips corresponding to customers with a fixed demand or skips which have already waited the maximum number of days) have not been served. In this case, the algorithm tries to serve these skips by means of two steps:

Step 1: Insert the skips into the constructed routes removing, if needed, skips with lower priority. The insertion is made by means of the cheapest insertion method while the removal is made by means of the maximum saving method. The insertions and the removals are made only if it is possible to satisfy the time windows of the customers and the plants whose time of service is changed. This procedure will decrease the penalty costs.

Step 2: If, after *step 1*, there are still unserved skips which have to be served during the day and there are drivers which have not yet been engaged for the service, this means that the algorithm can not find a skip corresponding to a customer which is open when the vehicle arrives. In this case, the algorithm chooses the skip for which the service requires the minimum time. Thus, the vehicle serves this skip even if it has to wait till the corresponding customer (or the plant, or both) opens. Then the algorithm returns to the beginning of *phase 2* to evaluate if there are still unserved skips which have to be served during the day.

Phase 2 ends when any of these three conditions is satisfied:

1. all the skips determined in *phase 1* are served;
2. there are no more drivers available;

3. all the skips which have to be served during the day are served and there is no customer in the list of the remaining requests which is open when the vehicle arrives.

Finally, in *phase 3* the algorithm tries to improve the solution found as follows.

1. Reduce the number of drivers engaged: if there are two routes that last much less than the shift-time, then the algorithm tries to put one of these routes after the other.
2. Serve skips with lower priority not considered in *phase 1*: if there are routes which last less than the shift-time, the algorithm tries to insert the requests not considered in *phase 1* into these routes till they cover the entire shift-time. The insertion is made by means of the cheapest insertion method and must satisfy the time windows of the customers and the plants whose service is delayed because of the insertion.
3. Reduce the cost of the extra time: if there are two routes which serve the same kind of skips such that one of them, say r_1, exceeds the shift-time and the other, say r_2, does not cover the entire shift-time, then the algorithm removes skips from r_1 and insert them in r_2. This procedure is repeated till r_1 does not exceed the shift-time or r_2 exceeds the shift-time.

A reduction of the number of drivers implies a reduction of the cost of the drivers engaged for the service but also a reduction of the number of skips served and, consequently, an increase of the penalty costs. Thus, the algorithm first evaluates the average penalty cost paid to delay the service of the skips served daily by a driver. If the cost of a driver is greater than the calculated penalty cost, then the algorithm executes *step 1* of *phase 3*, otherwise it goes directly to *step 2*.

At the end of *phase 3* the algorithm stops. More details on the algorithm can be found in Archetti and Speranza (2002). We have not proposed a more sophisticated algorithm (e.g., a metaheuristic) or an exact algorithm since we believe that the high complexity of the problem would make it ineffective. A simple algorithm, based on a local search, which takes into account all the characteristics of the problem seems to be the best choice in this case.

4 The Real Case Problem

The firm which carries out the service is a public firm located in Brescia, a city in the North of Italy. The customers are located in the county of Brescia, as we can see from Figure 1. In this figure the triangles represent the customers while the squares represent the plants. The depot of the vehicles is located in the city of Brescia and is represented by a circle. There is a plant which is not represented in the map: it is located in Milan, which is about 80 kms West of Brescia. The characteristics of the real case problem are the following:

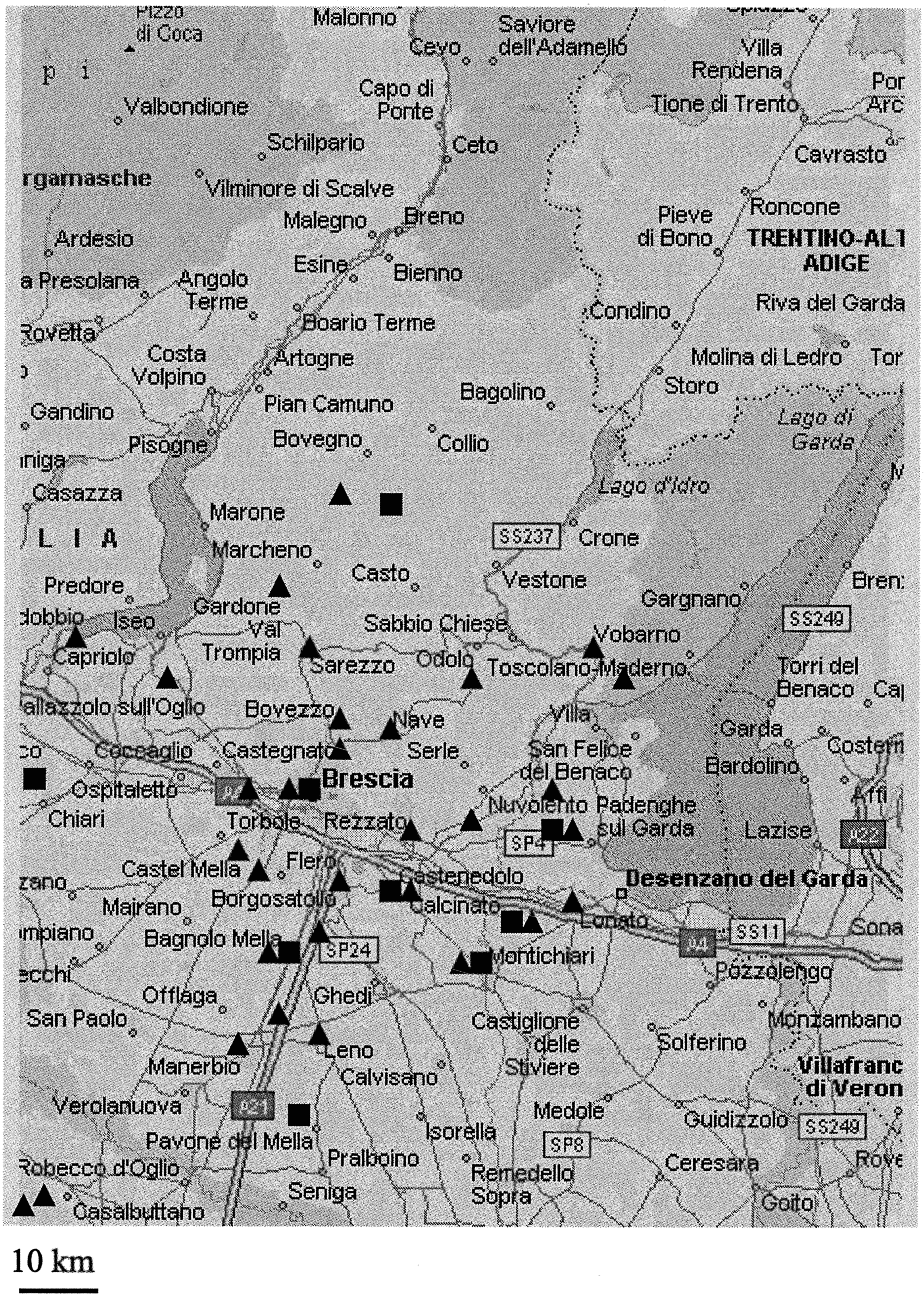


Fig 1. Customer and plant locations in the county of Brescia

1. **Number of drivers:** the number of drivers changes from day to day. The fixed cost of a driver is equal to the ordinary cost of the entire shift.
2. **Capacity constraint:** the capacity of each driver is one skip.
3. **Kind of waste:** there are 5 different kinds of waste.
4. **Shift-time:** each shift lasts 6 hours and the extra time cannot exceed 2 hours. There are two shifts in a day: the first is from 5:30 to 11:30 and the second is from 12:30 to 18:30. This implies that the algorithm plans the service twice per day: once for the first shift and once for the second shift. The hourly cost of the extra time is 30% greater than the ordinary cost.
5. **Kind of skips:** the skips have two different volumes: $10m^3$ and $20m^3$. For each volume there are two kinds of skips: the standard skips and the special skips. The special skips have to be transported back to the customer after being emptied and require the same service as the skips owned by the customers. Each kind of skip can transport any kind of waste. In practice, the skips of $10m^3$ are used for the organic waste while the skips of $20m^3$ are used for all the other kinds of waste.
6. **Time windows:** the time windows are different for each customer and plant.
7. **Priorities:** the customers with a fixed demand have to be served on the established day ($d_i = 0$), while the other customers have to be served within 2 days from the arrival of the request ($d_i = 2$).
8. **Service time:** each operation of loading and unloading a skip takes 15 minutes.

Because of the large number of constraints, in the real case problem it is a very hard task to find a feasible solution.

4.1 Computational Results

We tested the proposed algorithm on real case data for a period which goes from December 18, 2001 to January 5, 2002. The total number of working days during this period is 14 (the firm does not work on Sunday, on Christmas and on the New Year's Day). There are 13 plants where the skips can be emptied (each one accepts only one kind of waste) and 51 customers (including both the customers with a fixed demand and the customers with a variable demand).

We consider the requests from December 17, 2001 to January 3, 2002 as these are the requests served by the firm from December 18 (as the requests of December 17 cannot be served before December 18) to January 5 (on January 4 there was no request). The skips which are not served during a day are considered in the list of requests of the day after. The skips corresponding to the fixed demand requests are shown in Table 1. Table 2 reports the requests of service arrived during the period from the customers with a variable demand. The total number of skips which have to be served during the

Table 1. Number of skips corresponding to the fixed demand requests

Day	Number of skips
Monday	8
Tuesday	5
Wednesday	10
Thursday	6
Friday	13
Saturday	9

Table 2. List of daily requests

Day	Number of skips corresponding to the requests
17/12/2001	35
18/12/2001	32
19/12/2001	29
20/12/2001	38
21/12/2001	10
22/12/2001	16
24/12/2001	12
27/12/2001	13
28/12/2001	7
29/12/2001	9
02/01/2002	13
03/01/2002	6

period (including both the fixed demand requests and the variable demand requests) is 347. The total penalty cost is calculated summing up the daily penalty costs. The total number of drivers available during the period is 92. The average distance, measured in time, among the customers and the plants is 36 minutes.

For the evaluation of the solution cost, we have normalized the hourly labor ordinary cost to 1. Thus:

$c =$ hourly ordinary labor cost $= 1$;

$\bar{c} =$ hourly cost of the extra time $= 1.3$;

$c_v =$ cost of a driver $= 6$.

The solution applied by the firm is constructed by hand by an employee without any previous evaluation of the costs.

We have tested the algorithm for different values of c_{p_1} and c_{p_2}. These values are proportional to the daily cost of a driver with coefficients p_1 and p_2, respectively: $c_{p_1} = p_1 c_v$ and $c_{p_2} = p_2 c_v$. In the experiments, p_1 is always equal to 6% except for the case where $c_{p_1} = c_{p_2} = 0$. The reason is that, in the real instance, c_{p_1} is low and thus there is no reason to test larger values

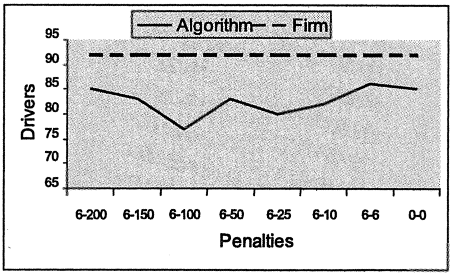


Fig 2. Number of drivers engaged

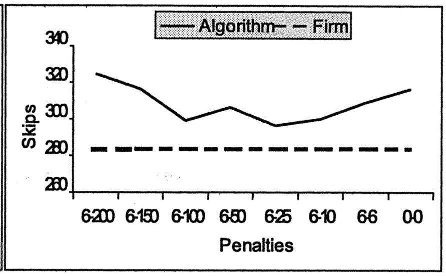


Fig 3. Number of skips served

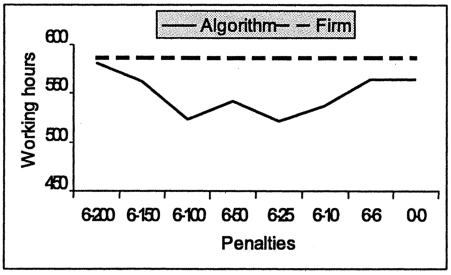


Fig 4. Extra time

of it. We, instead, vary p_2 between 0% and 200% since this cost can be large in the real instance. The running time of the algorithm is always lower than 5 seconds.

The solution applied by the firm uses 92 drivers to serve 284 skips with 33h55m of extra time. The computational results we obtained are shown in Figures 2-12. On the horizontal axis there are the pairs of values $p_1 - p_2$. From Figure 2 and 3 we can notice that all the solutions provided by the algorithm use a smaller number of drivers to serve a larger number of skips: for example, when $c_{p_1} = 6\%$ and $c_{p_2} = 100\%$ the algorithm uses 15 drivers less to serve 15 skips more. The extra time is instead greater than in the solution of the firm as we can see from Figure 4. When $p_2 = 200\%$ the extra time is 71h36m, i.e., more than the double of the extra time used by the firm. The data shows that the algorithm tends to reduce the extra time and the number of drivers used when the penalty cost decreases till $p_2 = 100\%$. The solution with parameter $p_2 = 100\%$ has the lowest number of drivers used. For lower values of p_2, the number of drivers increases again but the extra time decreases. Although the extra time used by the algorithm is larger, the total time is lower because the number of drivers used is smaller, as we can see from Figure 5. Figure 6 represents the total amount of working hours divided by the number of skips served, i.e., the average time to serve one skip. It is possible to see that the algorithm provides more efficient solutions than the one of the firm since it

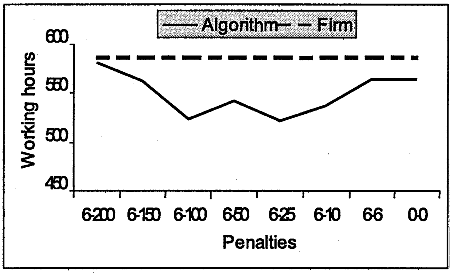


Fig 5. Working hours

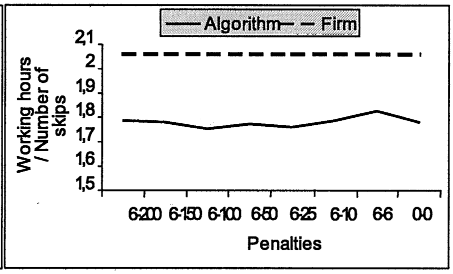


Fig 6. Working hours / Number of skips

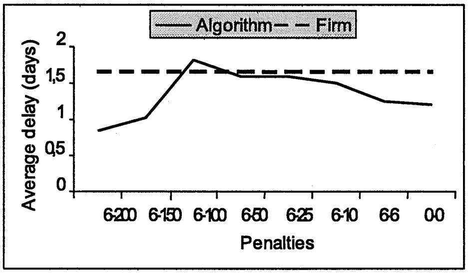


Fig 7. Average delay of the service

always uses less time to serve one skip. Another index of the efficiency of the solutions is given in Figure 7 which shows the average delay of the service, i.e., the average number of days each skip has to wait to be served. Except for the case where $p_2 = 100\%$, the algorithm provides solutions which reduce the delay with respect to the firm. The average delay increases when the cost of the penalties decreases till $p_2 = 100\%$ and then it decreases. The lowest value is 0.836 which corresponds to $p_2 = 200\%$. Figures 8 and 9 show, respectively, the average delay for the customers with a fixed demand and with a variable demand. From Figure 8 we can see that in the solutions of the algorithm the average delay of the customers with a fixed demand is always lower than the delay of the solution of the firm, whatever is the value of the penalty cost. The delay is very low when p_2 is high since the penalty cost paid in this case is very high. However, the delay increases when p_2 decreases with a maximum value of 1,00 when $p_1 = p_2 = 0$. The reason of this trend is that when p_2 decreases the difference between the penalty costs of the two groups of customers decreases and thus the delays of the two groups become more similar. From Figure 9 we can see instead that the delay of the customers with a variable demand is always lower than the delay of the solution of the firm except for the case where $p_2 = 100\%$ where it is equal to 2.498. The maximum delay of the customers with a fixed demand is always equal to two days except for the case where $p_2 = 200\%$ and $p_2 = 150\%$ where it is equal

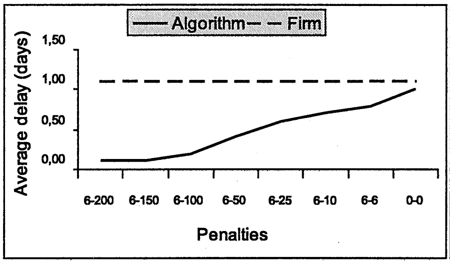


Fig 8. Average delay for the customers with fixed demand

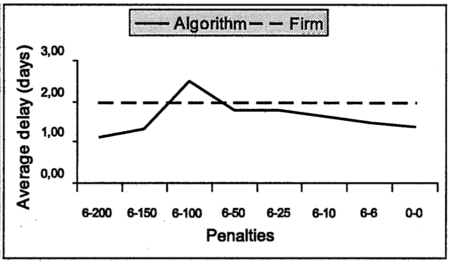


Fig 9. Average delay for the customers with variable demand

to one day. For the customers with a variable demand the maximum delay is always equal to five days except for the case where $p_2 = 100\%$ where it is equal to six days. In the solution of the firm, the maximum delay of the customers with a fixed demand is five days while the maximum delay of the customers with a variable demand is ten days. There is no relation between the number of skips of a customer and the corresponding delay, both in the solution of the algorithm and of the firm. Thus, from Figures 2-9, we can conclude that the algorithm gives more efficient solutions than the solution of the firm. It is possible to notice also that there is a trend to provide solutions more and more similar to the solution of the firm when the penalty cost decreases since the gap between the values of the different parameters tends to decrease. We now analyze the costs.

From Figure 10 we can see that the operational costs, i.e., the sum of the cost of the drivers and the cost of the extra time, are always lower than the costs of the firm (except for the case where $p_2 = 200\%$). Moreover, from Figure 12 we can see that there is a very large difference between the total cost of the solutions of the algorithm and the total cost of the solution of the firm for large values of p_2. This difference is explained by the penalty costs. Figure 11 shows that the penalty cost of the solution of the firm is much greater than the penalty costs of the solutions of the algorithm when p_2 is large. When the value of p_2 decreases, this gap drops down. We can also see that the trend of the costs of the penalties is the same as the trend of the total costs: the algorithm always provides solutions with lower costs, but for large values of the cost of the penalties the difference is large, while the gap tends to zero when the cost of the penalties decreases. There are three reasons that explain this large difference for large values of p_2:

- The solution of the firm serves a lower number of skips (in the case where $p_2 = 200\%$ the algorithm serves 14.44% more than the total number of skips served by the firm).
- The solution of the firm often delays customers with high priority for which the penalty cost is much larger.

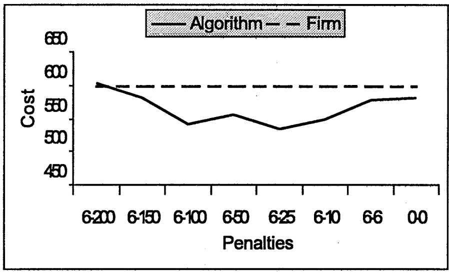


Fig 10. Cost of the drivers + cost of the extra time

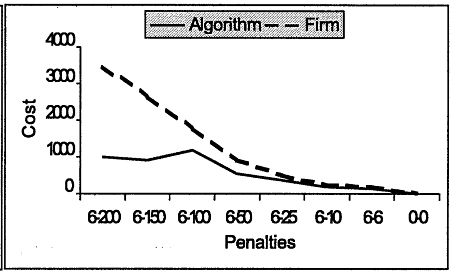


Fig 11. Cost of the penalties

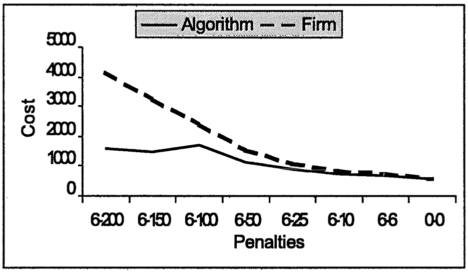


Fig 12. Total cost

- The firm prefers to reduce the cost of the drivers and of the extra time instead of the cost of the penalties. Note that also in this case the algorithm performs better than the solution of the firm as we can see from Table 3 where the results with $p_1 = p_2 = 0\%$ (i.e., without penalties) are shown. The number of drivers used is lower and, even if the extra time is larger, the total number of working hours is lower. The level of service to the customers is better since the algorithm serves 32 skips more (that is, 11.27% of the total number of skips served by the firm) with a lower average delay. Finally, the total cost is lower.

We can conclude that the solutions given by our algorithm are much better than the solution applied by the firm since the level of service to the customers is higher and the cost is always lower than the cost of the solution of the firm, even when we do not consider the penalties.

The algorithm has been implemented by the firm but it is not currently in use because of the difficulties met in the maintenance of the data bases needed by the algorithm.

Table 3. Comparison between the solution of the algorithm and the solution of the firm when $p_1 = p_2 = 0\,\%$

	Solution of the algorithm	Solution applied by the firm
Number of vehicles used	85	92
Extra time (hours)	53.45	33.917
Working hours	563.45	585.917
Number of skips served	316	284
Working hours / Number of skips	1.783	2.063
Average delay of the service	1.193	1.686
Total cost	579.485	596.092

Conclusions

We have considered a real case problem, the 1-skip collection problem, where a fleet of vehicles with a maximum capacity of one skip must collect a number of skips located at the customers. The skips contain waste of different kinds and have to be emptied in plants which accept the same kind of waste. Each time a vehicle collects a skip from a customer, it has to go to a plant and empty it. A number of constraints and problem characteristics have to be considered, as time windows for the customers and the plants, shift-time, different kinds of skips, number of drivers available to carry out the service and priorities assigned to the customers that have to be served. The objective is to minimize the total cost of the service given by the sum of:

- the cost of the drivers engaged to carry out the service;
- the cost of the extra time;
- the penalty costs paid for the unsatisfied requests.

We have designed a heuristic algorithm to solve the real case problem. Because of the large number of constraints, the algorithm first finds a feasible solution by means of the nearest neighbor algorithm, then improves it. The computational results show that the solutions of our algorithm are much better than the solution applied by the firm which carries out the service since a larger number of skips is served with a smaller number of drivers. The cost of the solutions of our algorithm is always lower than the cost of the solution of the firm. Even in the case where we do not consider the penalties the solution of the algorithm has a cost which is 97.21% of the cost of the solution of the firm and serves 32 skips more (that is, 11.27% of the total number of skips served by the firm).

Acknowledgements

We wish to thank Mario Tomasoni, the manager of the firm which carries out the service, for his support.

References

Archetti, C. / Speranza, M. G. (2002): Vehicle routing in the 1-skip collection problem. To appear on *Journal of the Operational Research Society.*

Ball, M. / Magnanti, T. L. / Monma, C. L. / Nemhauser, G. L. (1995): *Network Routing,* Handbooks in Operations Research and Management Science 8, North–Holland.

Ball, M. / Bodin, L. / Baldacci, R. / Mingozzi, A. (2000): The rollon–rolloff vehicle routing problem. *Transportation Science*, 34:271–288.

Cordeau, J-F. / Gendreau, M. / Laporte, G. / Potvin, J-Y. / Semet, F. (2002): A guide to vehicle routing heuristics. *Journal of the Operational Research Society*, 53:512–522.

De Meulemeester, L. / Laporte, G. / Louveaux, F. V. / Semet, F. (1997): Optimal scheduling of skip collections and deliveries. *Journal of the Operational Research Society*, 48:57–64.

Toth, P. / Vigo, D. (2002): *The Vehicle Routing Problem.* SIAM Monographs on Discrete Mathematics and Applications, Philadelphia.

Planning and Controlling Daily Deliveries to the Same Clients: Problems and Adequate Solutions

Thomas J. Bieding

Department of Economics and Business Administration, Bergische Universität Wuppertal, Germany

Abstract: Planning and controlling daily deliveries to the same customers involves numerous difficulties. One of the major recurring problems is inaccurate or incorrect relevant data. Using appropriate strategies still made it possible to successfully implement delivery planning solutions. By introducing an adequate and workable measurement system it is now feasible to collect more exact data about the delivery time at each customer to improve the results. With the help of this data we can validate the actual vehicle routing approach precisely for the first time. Finally the first realised planning project based on that data gave evidence that it is possible to increase the quality of planning and to realise higher savings.

1 Introduction: Problem Overview

In Germany every press whole seller has a sole right of distribution newspapers and magazines to the resellers in his area. Therefore the press whole seller must deliver the products from nearly every publisher to ever reseller. Especially the newspapers arrive late at night. This means that the delivery could not start before this point. On the other side the products have to be delivered before the opening times of the resellers. So the daily delivery of newspapers and magazines in Germany is a strategic vehicle routing problem with restricted loading capacities and time windows. For solving the vehicle routing problem for press wholesalers using the method of Dillmann (For detailed explanation of this method see Becker et al. (1994); Dillmann et al. (1996); Dillmann (1999) and Dillmann (2000). See Dillmann (2002) especially for explanation of the data problem in context to his planning method) it is necessary to have data referring to the delivery points and the routes. This includes specifically the realistic (average) starting time from the depot and a realistic (average) arrival time at the last customer. As a result we get a realistic (average) duration for each route. In connection with data from the delivery points (geographical co-ordinates, latest delivery time, delivery volume, route assignment, sequence within the route) a parameterisation takes place. With the aid of that parameterisation the new route assignments and the new sequences can be proved in reference to delivery in time, overall weight of each route et cetera. While the data referring to the delivery points already exists, the duration of each route has to be surveyed. For that purpose normally the shipping supervisor writes down the starting time during a fixed period and gets the arrival

B. Fleischmann et al. (eds.), *Distribution Logistics*

time at the last customer from the driver of each tour. It is obvious that reactive effects occur in that form of data collection. Either the driver will state the arrival time at the last customer very conservative (with additional buffer time) or his mode of operation will decelerate after arrival at the last critical customer (customer with the minimum buffer time) till the end of the tour. We discovered this behaviour in many cases.

Because of the fact that the tour duration is a very important value in connection with the parameterisation a fuzziness of the original data means fuzziness in modification of the distribution structure. We assume that the full potential for improvements and savings could not be tapped. According to experience the already existing data from the press wholesalers databases are deficient. That concerns the delivery sequences. Very often parts of the data did not represent the driving sequence in reality. So the already existing inaccuracy of the duration can be partly enlarged through wrong sequences. Figure 1 and 2 show an example for this case.

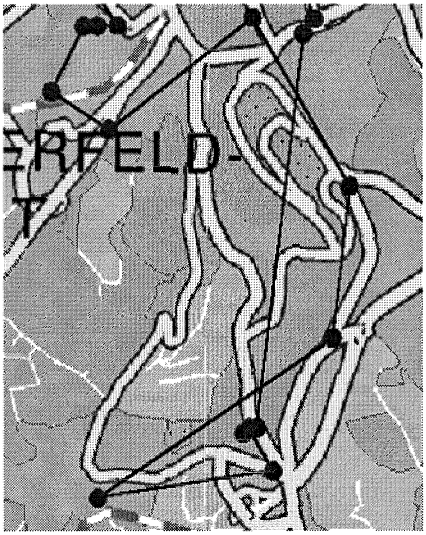


Fig. 1. Driving Sequence represented by the stored data

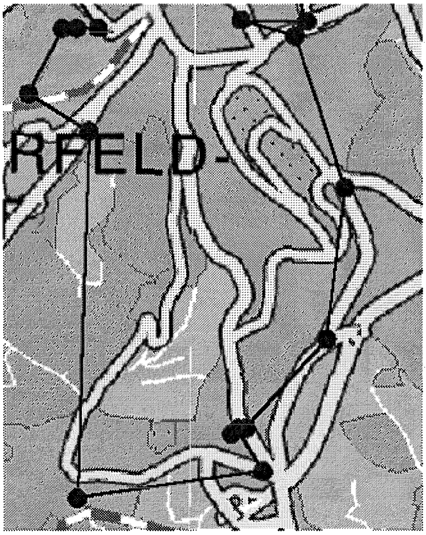


Fig. 2. Driving sequence in reality

The time amount for the sequence in figure 1 is higher than for the sequence shown in figure 2. If we use the tour duration for the model representation (Fig. 1) from the tour where the sequence in figure 2 is part from, we over-estimate the driving speed in our model.

Furthermore the stated latest delivery times could be mistrusted in reference to their correctness. We know it from several projects, observations of the daily practice and numerous discussions that the latest delivery times were very often asked in written form and stored without controlling them. Sometimes even a standard scheme is used: customers with weekly sales up to for e.g. 200 € have to be delivered 15 minutes before their opening time. Customers with more than 200

€ weekly sales even 30 minutes before. Such a procedure can tighten the underlying planning problem very much.

The significance of realistic tour starting times, which are result of the publishers delivery arrival time as well as from the picking performance, was underestimated sometimes. In one case the wholesaler confirmed our request to earlier starting times of some tours. After rearrangement of our plan all tours were picked in time and ready for departure. As the drivers of the early starting tours were also integrated in the picking process of the later starting tours, their tours could only depart with a delay between 20 and 30 minutes.

In another case we needed a tour starting time before the earliest stated time. We were informed that a supplier could not deliver earlier. We were very persistent because getting an earlier starting was very important to save one tour. We explained the need of that earlier starting time in detail and challenged the supplier's delivery time. Because of the discussions with that supplier we finally could identify solution strategies to realise an earlier delivery. In conjunction with speeding up the preparations for the picking process we could realise our required starting time for one tour.

Of course in practice we very often come across with structural defects of planning problems. As others too we have found strategies to come along with the described problems so that we can realise satisfying results anyhow. (For examples of that see Dillmann (2002). See also Bauer (2002)). Through our experiences with several consulting projects with different data quality we were convinced that a higher quality of data will lead to much better planning results. So we focussed our research on a practically feasible system, which could permanently log the delivery time and help to determine the duration of each delivery tour and the sequences very easily. Additionally such a system could make it possible to verify our parameterisation and the realised results of the planning process as well as whether the stated latest delivery times are really relevant or not. We were convinced that this would help to get a high level of transparency about the daily deliveries, which has not existed yet. In the following we describe in the second part two possible measurement technologies (GPS and RFID) in detail with their assets and drawbacks to explain our decision for transponder based measurement. In the third chapter we show how we use the collected data to establish a controlling system for daily work. Furthermore this helped us to validate the used planning approach. The impact on data quality and optimisation work is explained in chapter five as well as positive internal and external effects of the used technology. Finally in the last chapter we evaluate the results to deduce recommendations for the practical work as well as important questions for further scientific works.

2 Logging of Delivery Times

We identified continuous logging and logging on the spot as possible starting points to get valid data about the driving sequences of each tour and the daily delivery time at each customer.

2.1 GPS based data logging

A continuous logging of the track could only be realised with the help of GPS technology. (For details about the GPS technology see e.g. Hoffmann-Wellenhof/Collins (2001). See also Mansfeld (2001)). In cooperation with a manufacturer of GPS devices we modified an existing device according to our requirements and tested a prototype. The device was improved based on the experiences gained with the prototype. As a result we had a very small data logger powered by rechargeable batteries and an additional power supply connect to a car's power outlet. While turned on, the device stores geographical co-ordinates with a timestamp. By experience, we identified the following logging parameters to achieve good results:

- Minimum logging interval 1 second
- Maximum logging interval 20 seconds
- Threshold value for shift of position 20 m

If there is no shift of position the co-ordinates will be stored every 20 seconds together with time and date. If the shift of position is more than 20 metres co-ordinates will be stored at most once per second. After the tour has finished the device will be connected with a serial interface to a computer and the data will be downloaded. The data than will be imported into a Geographical Information System (GIS) for further analysis. Here the data can be visualised, edited and analysed. Editing the raw data is only necessary if the logging process has started before the driver starts his tour or if the process finishes after the tour has been completed.

As a result of processing, the system delivers the exact starting time based on the atomic clocks of the GPS satellites, arrival time at the last client and the amount of kilometres actually driven. The visualisation on maps makes it possible to reconstruct the drivers route (see figure 3).

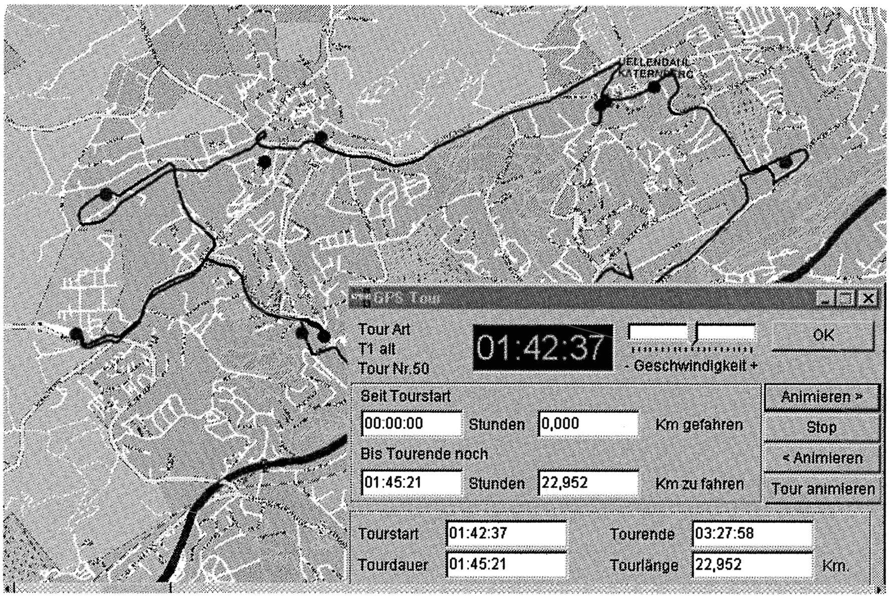


Fig. 3. Visualisation of a GPS Track in a Geographic Information System (GIS)

Our first attempts to calculate the arrival time with the help of the GPS data revealed the problems of that policy. The analysing routine identifies every shift of position of less then 1 meter between two successive measuring points as a stop. After that the corresponding time will be assigned to the nearest client (see figure 4).

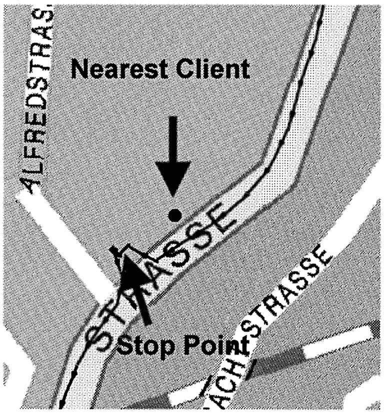


Fig. 4. Nearest client to a GPS stop

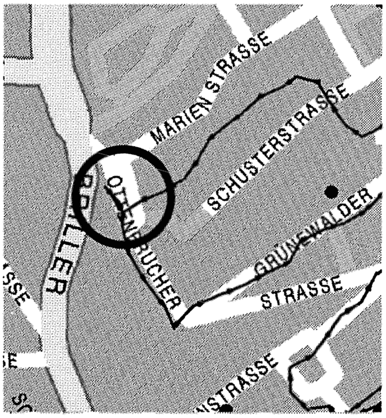


Fig. 5. GPS Stop without client

In fact, a stop can be identified without problems but the explicit and correct assignment of this stop to a client co-ordinate is sometimes problematic. It may happen that a driver stops without delivering any client (see figure 5). As such a stop could be as long as a delivery stop, an automatic classification of the stops to prevent such problems seems to be impossible. Missing client co-ordinates have a comparable impact. So in both cases wrong assignments and consequently wrong recorded delivery times can appear.

Another problem lies in wrong client co-ordinates, which cause the same problem. Using the delivery sequences could theoretically help to solve some of the problems but in reality the drivers very often change the sequence.

Using a maximum range for each stop in which the routine searches for a client could be a solution for some cases. But in other cases it could lead to the same problem. What happens when the driver stops to have a coffee break and there are clients within the range which will be served later? Additionally high buildings, deep valleys or forests can cause interference's with reception.

Independently of our efforts, other wholesalers who tested GPS technology for delivery time measurement made the same experiences. Based on the described problems we abstained from using GPS for delivery time measurement. But for detailed analysis of the run we recommend using GPS as a helpful technology.

The following Table 1 summarises the advantages and disadvantages of a GPS based measurement system.

Table 1. Pro's and con's of a GPS based delivery time measurement system

Pro	Con
Ready to use system	High Price: approx. € 900 for each tour
Preparation of delivery points not needed	Co-ordinates of every delivery point needed
Recording of the complete tour	GIS system needed for analysing data
Information about the driven kilometres	Daily data handling and device preparation takes a lot of time
Information about the driven way	Raw data has to be edited
Can be combined with GSM modem for data transmission	Distinct assignment of GPS stop to client coordinate not always possible
	Data transmission over GSM-networks causes additional costs

2.2 Transponder based delivery time measurement

For measurement on the spot the barcode technology seemed to be appropriate. Every point of sale could be equipped with a barcode label that would be scanned during the delivery process and stored in a corresponding device. But as an optical method we considered this to be very error-prone because water, dirt and/or damage of the label could disturb the scanning process. So we evaluated this as a not practicable and not problem-adequate solution. As a robust alternative we chose the transponder technology and developed a solution based on radio frequency identification (RFID). (For details about the RFID technology see e.g. Finkenzeller (2003) and Paret (2002). For a survey about identification technology in logistical systems see Pflaum (2001)). A RFID Systems consist of two components: an either mobile or stationary reader unit and an information carrier called transponder or tag. Because of the actual requirements, we chose a passive (without own power supply unit) read-only system. As the transmission of the information is based on electromagnetic waves, the system works through liquid, gaseous and solid materials. To read out the transponder information, the corresponding scanner builds up an electromagnetic field, which activates a microprocessor on the transponder to send back its information which is received by the Scanner. It will be decoded and stored. For identification we use an unique ID that is stored in each transponder during construction. We use a very handy device that only has one operation button and a display (see figure 6).

The maximum scanning distance between scanner and tag is 100 mm. Because of this technical specification we can guarantee that the driver was at the place where the transponder is fixed. To further simplify this easy handling, the use of wearable ring scanning systems for hands-free measurement can be implemented. Unfortunately, the systems already are use for barcode scanning are not yet available for the RFID technology. The transmission of the stored data from the scanner to a computer is provided by an infrared interface.

As preparation for the first installation (The case study is described in Bieding (2001)) all the transponder ID's had to be assigned to single clients. After that, the placement of the tags started. First tests started in spring 2001 and already in summer we finished the area wide installation for a customer with nearly 1000 delivery points. At the end of the year we had comprehensive and valid data about the actually starting and delivery times.

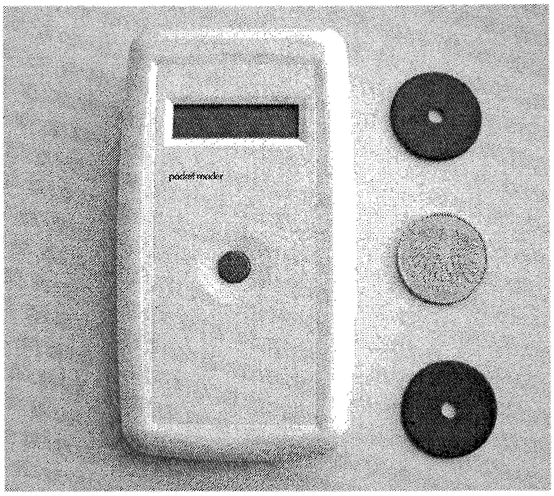


Fig. 6. RFID Reader and Tags

As expected, the scanning rate was not 100%. The reasons were lost transponder, unknown location of the transponder, temporary driving personnel who did not know the systems, driver who just forgot scanning and sometimes damaged devices. The loss was mainly caused by bad fixation. To avoid this we gave explicit instructions how to fix the transponder.

If the driver does not know the position of the tag, he cannot scan it. So detailed information about the placement is very important to realise a high scan rate. We recommend that the driver is responsible for the placement and that this information is stored at a central place. Changes of the tag placement have to be communicated and documented. So whenever drivers change the tour or temporary driver personnel is employed, detailed information about the placement is available.

New drivers have to be instructed on how to use the system. Of course it can happen that a driver forgets to read a transponder. Sometimes the driver just forgets the scanner in his car or to scan a tag because he is in hurry.

Naturally the device could be damaged due to general technical defect or wrong handling. These cases could never be ruled out completely. But not scanned transponder can be avoided through a quick analysis of the data and of course feedback to the driver.

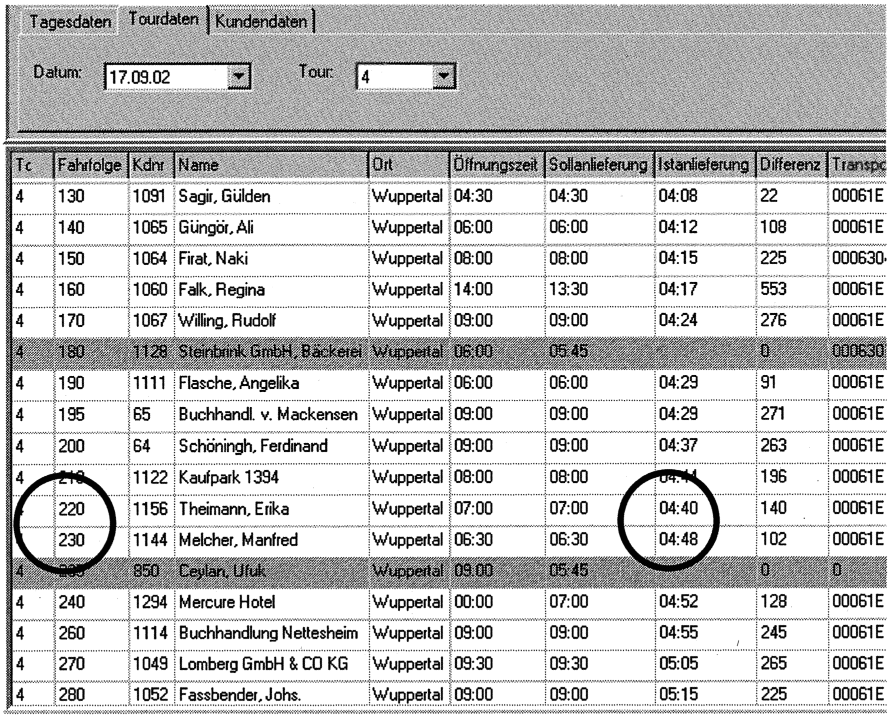
According to the previous chapter, the following table 2 summarises the advantages and disadvantages of a transponder based delivery time measurement system.

Table 2. Pro's and con's of transponder based delivery time measurement

Pro	Con
Easy to use system	Every delivery point has to be equipped with a transponder
Low price: under € 500 for each tour	Records only delivery time
Distinct assignment of transponder time stamp to delivery point	Data transmission over GSM-networks causes additional costs
Easy data handling	Tranponder can get lost
Can be combined with GSM modem for data transmission	Driver needs detailed information about the tranponder placement
Robust components	
Valid data for controlling purpose	

3 Tour Controlling

For the download process from scanner to computer we have developed an easy to use software tool, which stores the data in files or directly into a database. For about 50 records the process is finished after almost 15 seconds. It is much faster than the complete procedure for a GPS device. Our analysing routine, which is completely integrated into the existing ERP-Tool, allows the user to analyse the data directly after the read out process finishes. The user can choose between a reporting by days, customers or routes or a combination of all. The major advantage of the integration into existing applications is the access to all relevant data like client ID, latest delivery times, tour assignment and sequence. The data delivered by the transponder is associated with the customer. Then, the buffer time (difference between expected time and actual delivery time) is calculated and all missing delivery times are highlighted (see figure 7). In the present case two transponders were not scanned. Now the user is able to question the driver about the missing values to prevent such situations in future. Furthermore it is evident that sequence 220 (04:40) was delivered before sequence 210 (04:44). Analysing the past delivery days showed that it was not an exception and the stored sequences have to be corrected.



Tc	Fahrfolge	Kdnr	Name	Ort	Öffnungszeit	Sollanlieferung	Istanlieferung	Differenz	Transpc
4	130	1091	Sagir, Gülden	Wuppertal	04:30	04:30	04:08	22	00061E
4	140	1065	Güngör, Ali	Wuppertal	06:00	06:00	04:12	108	00061E
4	150	1064	Firat, Naki	Wuppertal	08:00	08:00	04:15	225	000630
4	160	1060	Falk, Regina	Wuppertal	14:00	13:30	04:17	553	00061E
4	170	1067	Willing, Rudolf	Wuppertal	09:00	09:00	04:24	276	00061E
4	180	1128	Steinbrink GmbH, Bäckerei	Wuppertal	06:00	05:45		0	000630
4	190	1111	Flasche, Angelika	Wuppertal	06:00	06:00	04:29	91	00061E
4	195	65	Buchhandl. v. Mackensen	Wuppertal	09:00	09:00	04:29	271	00061E
4	200	64	Schöningh, Ferdinand	Wuppertal	09:00	09:00	04:37	263	00061E
4	210	1122	Kaufpark 1394	Wuppertal	08:00	08:00	04:44	196	00061E
4	220	1156	Theimann, Erika	Wuppertal	07:00	07:00	04:40	140	00061E
4	230	1144	Melcher, Manfred	Wuppertal	06:30	06:30	04:48	102	00061E
4	[illegible]	950	Ceylan, Ufuk	Wuppertal	09:00	05:45		0	0
4	240	1294	Mercure Hotel	Wuppertal	00:00	07:00	04:52	128	00061E
4	260	1114	Buchhandlung Nettesheim	Wuppertal	09:00	09:00	04:55	245	00061E
4	270	1049	Lomberg GmbH & CO KG	Wuppertal	09:30	09:30	05:05	265	00061E
4	280	1052	Fassbender, Johs.	Wuppertal	09:00	09:00	05:15	225	00061E

Fig. 7. Tour wise analysis of delivery time

Another application (see figure 8) allows an assessment at one glance whether the complete delivery was on time or not. For this visualisation, all planned service times are sorted in increasing order with their corresponding actual service times. The horizontal-axis shows the clients and the vertical-axis the time in minutes after midnight.

Whenever the actual time curve (upper curve) intersects the reference time curve (lower curve) there is a late delivery. This form of visualisation is perfectly suitable for single tours. (see figure 9). Here is clearly to see that almost 4 clients are time-critical. In comparison to another delivery day (figure 10) it is obvious that the performance decreases at the end.

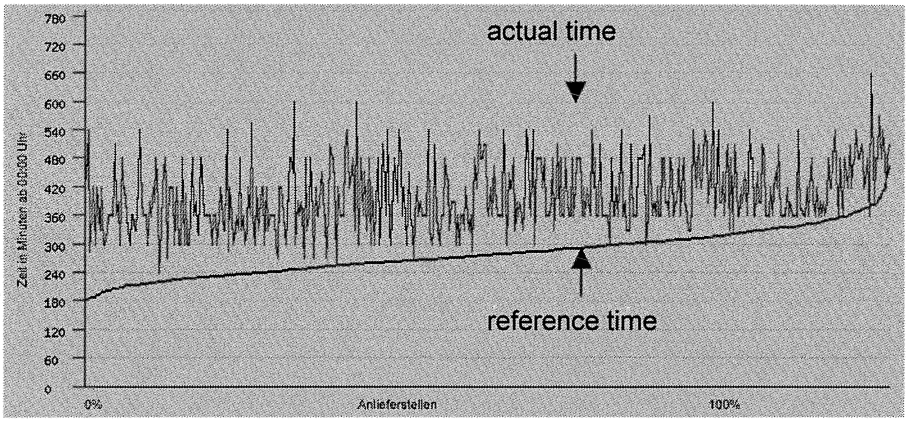


Fig. 8. Visualisation of all measured Service times (actual time) in proportion to their latest service times (reference time) of a delivery day

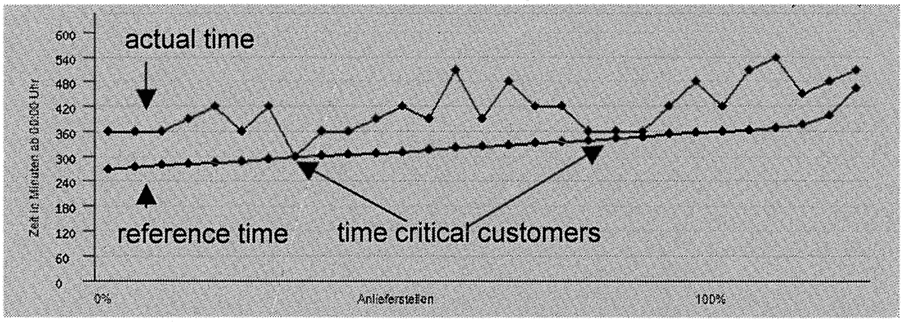


Fig. 9. Real service times (actual time) in comparison to latest service time (reference time) of a tour. Decreasing Performance

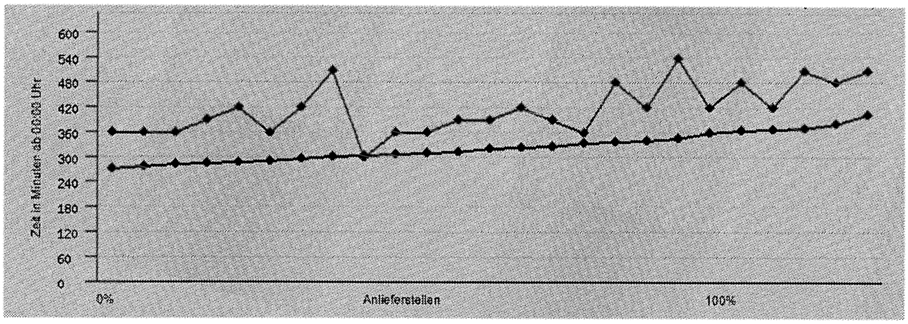


Fig. 10. Real service times (actual time) in comparison to latest service time (reference time) of a tour. Normal Performance

As can be seen in figure 10, the sequences were changed (less time critical customers) and 4 delivery points were assigned to another tour.

For a daily (tactical) controlling it would be desirable to consolidate all information to one index, which allows evaluating the delivery quality. We believe, that because of the complexity of the delivery structure this is not feasible. (See Kopfer (2000) for a similar assessment about the complexity of contemporary logistic systems.)

We recommend to monitor the following aspects:

- Delivery in time (all tours, single tour, customer)
- Delivery sequence (are there changes)
- Starting time (did the tour started on time)

If there are particular deviations they should be researched in detail. In those cases the tactical controlling can be expanded. For this purpose we need to define the values "estimated tour duration", "tour buffer time" and "estimated tour buffer time".

The **estimated tour duration (ETD)** is calculated the following way:

$$ETD = S_{\varnothing} * d + t_s * n_s + ut_p * n_p$$

Where:

$S_{\varnothing}$	= average Speed
d	= distance
t_s	= stationary time
n_s	= number of stops
ut_p	= unload time per Parcel
n_p	= number of parcels

Tour buffer time is the minimum difference between actual time and reference (latest delivery) time of each customer within a tour.

The **estimated tour buffer time** is the minimum difference between the estimated arrival time and the reference (latest delivery) time of each customer within a tour.

Based on a detailed survey we additionally propose to use the following values for evaluating the delivery service. They could be seen as a strategic enhancement.

- Deviation between actual starting time/reference starting time and real tour duration/ estimated tour duration,
- number of delayed deliveries and
- deviation between real tour buffer/estimated tour buffer

Table 3 shows such a controlling list for a single delivery day.

Table 3. Example for a daily tour controlling

Tour	Estimated tour duration	Real tour duration	Difference in minutes	Difference Starting Time	Delayed Clients	Tour buffer differences
1	01:31	02:09	-38	-19	0	0
2	01:45	01:45	0	+29	1	-29
3	02:04	02:18	-14	-5	0	+5
4	01:47	01:44	+3	+1	0	-10
5	01:51	02:08	-17	-11	1	-7
6	02:16	02:15	+1	+24	6	-23
7	02:19	02:12	+7	+12	0	+3
8	01:29	01:29	0	+19	1	-17
9	02:23	02:11	+12	+26	2	-8
10	01:54	02:40	-46	-11	0	+14
11	02:34	01:35	+59	-10	0	+37
12	02:04	01:55	+9	-17	0	+3
13	02:42	02:30	+12	-17	0	0
14	02:32	02:13	+19	+10	0	+10
15	02:26	02:40	-14	+10	0	-12
16	01:50	02:07	-17	-30	0	-30
17	02:08	02:10	-2	+3	0	+20
18	02:03	02:49	-45	+10	0	+10

In this example Tour 1 started 19 minutes earlier, all clients were delivered on time and the deviation between actual and estimated buffer is 0. Still the actual tour duration differs about 38 minutes from the estimated duration. This triggers an detail analysis of this tour. In this specific case the tour started 19 minutes earlier than estimated. But the driver arrived on estimated time at the point of delivery. Up to this point, the driver needed 19 minutes more but the route buffer time (minimum difference between actual starting time and reference time) did not change because he started earlier. Subsequently there was no time-critical client anymore. This missing pressure explains the decreasing speed. The driver just took his time. Because the objective to realise a delivery on time was met, the longer overall duration of the tour is no issue for the model.

For tour 2 the calculated and real duration is the same. The client delivered with a delay of 29 minutes is explained by the about 29 minutes delayed starting time. The reason for this delay needs to be identified.

The deviation of tour 10 is conspicuous. It took 46 minutes longer, but none of the clients had a delayed delivery. Analysis of the tour shows that the driver has changed the default sequence. The critical client originally at position 32 within the tour was already served at position 24. Additional changes from the original sequence caused the increased tour buffer time. The duration also increased because of the longer distance to drive. That means a new calculation of the tour is necessary using the changed sequences.

The extremely shorter duration of tour 11 is caused by the fact that this driver delivers only the time-critical newspapers during the night. Most of the non time critical but time consuming magazines were already delivered in a pre-delivery tour. Additionally the driver can now take the shortest path because he has won a lot of time. The calculations have to be updated with the new weights and sequences.

Those cases should illustrate the complexity of this specific distribution structure and show that it is impossible to reduce it to one single index. But the above explained controlling approach offers stepwise problem adequate utilities to research specific problems in detail to get deeper insight and to identify adequate solution strategies. Nevertheless this requires staff members understanding the underlying problem.

Finally this shows that the planning approach used is sufficiently close to reality. For several wholesalers we have analysed the daily delivery time in comparison with our estimated delivery time. In most cases our estimation had only a deviation of -/+ 15 minutes to the actual delivery times.

4 Use of Delivery Time Measurement

Of course, the question arises about the benefit of delivery time measurement. The answer is mainly given by the positive impact on planning and data quality. But we also see internal and external side effects.

4.1 Impact on planning and data quality

As we have shown it is possible to establish a daily delivery controlling by using the transponder-based delivery time measurement. For a vehicle routing project, we based our calculations used this data collected by transponders. Using the raw data we tried to identified realistic start and end times for every tour. To fulfil this task, data was cleared of non-recurring, specific situations which led to not reproducible results: Delayed deliveries from the publishers, delays caused by extreme weather conditions, damaged delivery cars or a driver who had overslept. Those cases must not be considered for a vehicle routing. As a second step we have reconstructed the existing structure (sequences) in our planning system and did a parameterisation (for details see Dillmann (2002) pp. 293).

We have identified two different parameter sets for the relevant distribution area: one parameter set representing a normal and one representing a slower performance. Both reflect the characteristic of the delivery structure. In our planning approach it is important that we have one valid set of parameters for the complete structure. So we had to analyse every tour in detail. We found the slower tours either less time critical than the faster ones, or the slower tours had their critical points of delivery, where they had to regard earliest delivery times, at the end of the tour. So missing time pressure and/or idle waiting time at the clients caused unnecessary long tour durations. Based on the results of our analysis we decided to use the set of parameters representing the higher performance (faster) for the planning process.

For planning we used the well known approach from Dillmann with only one difference. We did not use the vehicle capacity as committing constraint which means that the planning was focussed on delivery on time, shortest path and the minimum possible vehicles. Not considering the capacities takes up the proposal of Dillmann (Dillmann 2000) to expand a pre-delivery concept when the time windows are getting smaller. This policy was possible because of more exact data about the single delivery weights and the duration. Pre-delivery tours are necessary to transport those capacities which are available before the time-critical night delivery to avoid a capacity overload. Thus the time-critical night delivery is significant relieved. It is not necessary to drive pre-delivery tours every day. But we had a case where the driver did the pre-delivery and the night delivery every day (see table 3 tour 11). Thus this driver was over 50 minutes faster than originally expected.

The relevance of pre-delivery tours is shown by figures 10 and 11. Figure 10 is an example for the distribution of the complete delivery weight for two successive weeks (white and grey) sorted by day. Aligning the planning to the maximum weight would have lead to tours with low utilisation at most days. Viewing the distribution only of the time-critical newspaper weight over the same period (figure 11) illustrates this.

To reach the highest possible savings, we have deviated from the time constraint in for a few tours because customers with low sales volume would have caused higher costs if we had delivered them in time or later clients would have been delivered with a delay. This strategy is supported by the results of a first survey about the impact of delays on the sales of the time critical newspaper ‚BILD' (Wolfertz 2002). Here it was analysed how far significant mean differences appeared between the sales in time and delayed delivery. Over a period of 60 weeks over 180.00 cases were regarded.

In fact there was a number customers with delayed deliveries and significantly decreasing sales. But on the other side there were delayed delivered customers with significantly increasing sales. That means there have to be customers delivered in time with significant decreasing sales. Overall there were significant decreasing sales at 5,3% of all delayed delivered clients independent of the degree of delay as well as at 2% of all cases (on time and delayed delivered clients).

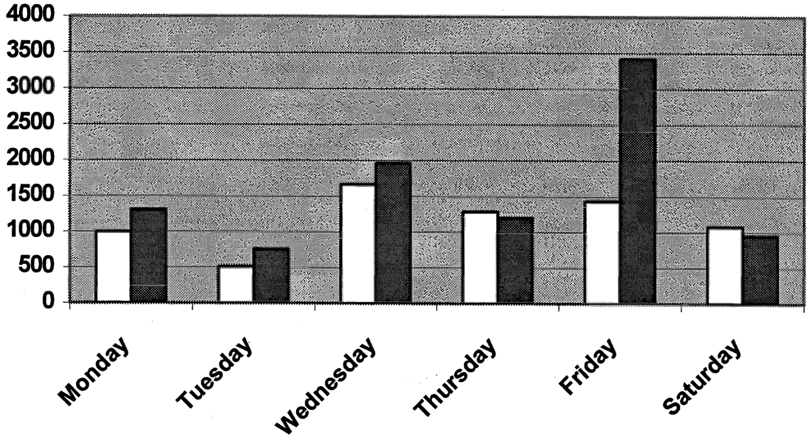


Fig. 11. Daily distribution of the complete delivery weight in kilograms for two following weeks of one tour

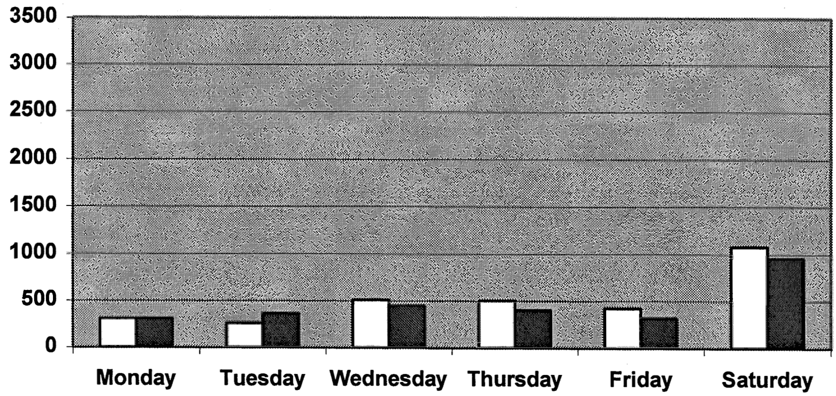


Fig. 12. Daily distribution of the newspaper delivery weight in kilograms for two following weeks of one tour

Most of the delays had no significant decreasing sales. On the other side there were significant increasing sales at 5,1% of all delayed delivered customers as well as at 2,3% of all cases. Based on these results it is impossible to deduce a correlation between delayed delivery and decreasing sales. Analysing the statistical correlation between delivery time and sales also gave no indication of a statistical correlation between sales and delay. Although some customers with delayed sell less, but whether this is caused by a delay is not checkable with the existing data.

A closer look to single cases gave explicit hints that there are probably other reasons. Once e.g. a shop was not delivered on time at the first day after a longer vacation. The decrease of sales was really significant. But it is absolutely normal that a shop needs some time after a vacation to get back to his normal sales because consumer needs some time to change their buying patterns. Analysing the sales of this shop after other vacation periods shows that this shop has similar decreases without delayed deliveries. Finally we can state the following conclusion:

- It is not possible to make the general assumption that delayed deliveries cause significant decrease sales.

In fact those results show that we have to view delays in a more differentiated way. The latest delivery times mentioned by shop owners or recorded by press wholesalers' employees could contain a dispensable buffer. In single cases it could be very helpful for the planning process to examine those data and to accept formal delays if it is useful.

Finally the information content of the measured data leads to a sophisticated planning approach, which builds the fundament for significant savings.

As a result we had a distribution structure and a corresponding payment structure, which saves 20% of the tours compared with the initial situation. Through successful realisation of that routing about 20% of the total annual distribution costs could be saved. For that purpose we have calculated a theoretically fair payment for the initial situation with help of the linear regression. As a result we obtained units of money for each driven kilometre, each client stop and each unit of weight. We have compared the new payment with the existing one tour by tour and could identify many differences. In single cases the payment was to high in other cases it was to low. This payment parameters were used to calculate the payment of each new tour. The first part of the overall saving was caused by the reduced total amount of driven kilometres. Additionally the press whole seller discounted the overall payment. After that we have distributed the tours to drivers in a way, that no driver had less payment than before. This is very important because drivers do not accept a lower payment but moderate increasing amount of work (more kilometres, more weight and/or more stops).

In comparable projects we realised in the past, where we used the same payment model, we could only save in average about 10% of tours and costs. As in the existing case, there were route plans we had to update (re-planning) some years after a complete new planning. This is necessary because of a permanent fluctuation of points of sale and changing latest delivery times.

As it is well known that the complete industry of publishers and press whole sellers is afflicted with massive decreasing sales (Pressegrosso 2002) the legitimate question comes up whether the reduced number of tours and the observed savings are mainly caused by the decreasing delivery volume or not.

For that purpose we analysed the number of customers from the mid of 2001 to 2002 and merely ascertained an absolute decline of 4 customers. Further we observed the delivery weight of every single tour from January 2001 to June 2002. It showed that the complete weight declined about 430t. This means about 16,5t lower weight each week or uniformly distributed a reduction of about 2.3t each day. During the planning process we have calculated with a time amount of 7.20 minutes for unloading 100kg. 100kg are about 12 magazine packages. A driver can carry a minimum of 3 or 4 packages at once. That means he has to go a maximum of 4 times from his car to the shop an back during this time. That is a realistic assumption which we have verified empirically. 7:20 minutes or 420 seconds multiplied with 23 (23 * 100 Kg = 2,3to) makes 9660 seconds or 161 minutes or 2:40 hours. This time saving is similar to the duration of a tour. So we had saved the time amount for one tour and theoretically we could see this as one saved tour. But we do not know how the weight and unloading time reduction is distributed over the delivery area of that press wholesaler. Furthermore we do not now the additional driving time is needed for the remaining tours to reach the clients from the theoretically obsolete tour. We know from several consulting projects that the time amount only for driving is not less than 1 hour. So we see the theoretically achievable time asset is in real life not strong enough to save a complete tour. Therefore we can attribute the realised reduction of delivery tours by three to the higher data quality respectively to deductions gained from experience and improved approaches for the intrinsic planning. We regard our results as strong evidence that the use of more exact data in connection with the vehicle routing for press wholesalers can yield significantly improved results Future projects will show whether these conclusions are correct.

4.2 External Effects

In the past a major nation-wide newspaper publisher urged his whole-salers to increase the transparency of the daily deliveries. The significance of that demand is made clear by the fact that the wholesaler which first introduced that system won a connotatively award. The quality of the daily delivery is now more transparent because information about it is available via Internet both for publishers and customers. Additionally the carrier and/or driver as a service provider for the wholesaler gets a valid proof of his delivery quality and is protected against unjustified complaints. Meanwhile several press whole sellers use delivery time measurement systems.

4.3 Internal Effects

The introduced measurement system makes it possible to establish a daily tour controlling. Delayed deliveries to customers, not scanned transponder and deviation from the given sequences can be easily identified.

A permanent measurement of relevant data like start time, duration and sequences of each tour takes place. Thereby and in combination with a strategically controlling exact data and, as a result, adjusted planning models are available for a fast reaction on changed delivery structures, ad-hoc solutions for issues and customer support and of course for a complete re-planning. The availability of reliable and current information about the deliveries can be valuable asset when dealing with customers. In one actual case a customer complained about a delivery that did not took place. The responsible employee could prove that the delivery already took place. As the customer got delivered before the opening time the products were unloaded not at the usual place but at the place for early deliveries. After getting this information the customer found the products there and excused for his complaint.

5 Conclusion and Future Work

Although, measuring the service time at the point of delivery based on RFID technology seems to be not as much cutting edge approach as using GPS, it is much more cheaper, easier to use and delivers absolutely reliable data day by day. The corresponding software technology reduces the post processing load for the daily controlling to a minimum and the person in charge has only to take care about small number of variables. We have introduced a set of values which indicate if there were any problems during the completed deliveries or if we have an incorrect model representation of our planning approach .

In our study the content of the measured data led to a sophisticated planning approach. Analysing the delivery times showed that there was considerable potential for optimisation. Another result was it is preferable to minimise the overall buffer time in a tour in order to decrease the overall duration, although this makes a tour vulnerable to delays when the starting time could not be met.

We did not use vehicle capacities as restrictions and in some cases we eased the time restrictions when it led to a useful solution. We found out that a tour gets significant faster when only time-critical newspapers are delivered. This makes it possible to increase the number of served clients through a single tour. On the other side this make it necessary to have pre-delivery tours for the non-time-critical magazines.

So the objective was to find an 'optimum' between pre-delivery and main delivery, which could lead to minimum feasible transportation costs. In comparison to past projects under similar conditions we could duplicate the annual saving. The return on investment into this technology was under 3 month. Additionally, we were able to show that the underlying latest delivery times were not necessary relevant for the sales in all cases.

For the future it seems to be necessary to verify the use of delivery time measurement. The refinement of the Dillmann planning approach points out some aspects which were not deeply researched until now:

- Analysing the impact of a stepwise disintegration of combined delivery for press wholesalers. (E.g. what are the economical effects if big time critical newspapers (BILD newspaper) get exclusively delivered ? Is it possible to separate the magazine distribution from the newspaper distribution in general?)
- Identification of corresponding decision rules (at which circumstances it is right to separate the distribution in which way?)
- Analysing the impact of conscious violations against the reference time restrictions more in detail. (Are the sales decreasing when some customers will be delivered delayed? What savings could be realised when some customers get a delayed delivery?)
- Finding strategies to get the real and therefore relevant latest delivery time.
- Analysing the effects of complete or part wise combination of the distribution structure of two or more logistics provider in the same region. (E.g. co-operation between press whole seller and regional newspaper distribution or pharmaceutical whole sellers)

Additionally the technological progress offers new opportunities for delivery time measurement which have to be analysed. In the meantime it is possible to use standard cellular phones as easy internet clients or to run customized applets. The driver could get a complete routing list on this device and is able to confirm every delivery with his cellular phone. In combination with location based services it is possible to verify the drivers location while he is confirming a delivery. Especially for nationwide distribution systems this could be an adequate solution.

Getting deeper insights about the relevant influencing variables on time-critical distribution logistics will possibly help to find more efficient and contemporary solution strategies.

References

Bauer, A. (2002): The Practice of Distribution Network Planning: Coping with Shortcomings in Important Data Quality. In: Klose, A. / Speranza, M. G. / van Wassenhove, L: N. (eds.): Quantitative Approaches to Distribution Logistics and Supply Chain Management. Springer Verlag, Berlin.

Becker, B. / Beckefeld, V. / Dillmann, R. (1994): Tourenplanung im Pressegrosso. In: Dehrigs, U. / Bachem, A. / Drexl, A. (eds.): Operations Research Proceedings, pp. 547-522. Springer-Verlag, Berlin.

Bieding, T. J. (2001): Wann wird der Kunde wirklich beliefert? In: *Logistik Heute* 6/2001, pp. Huss-Verlag, München.

Dillmann, R. / Becker, B. / Beckefeld, V. (1996): Practical aspects of route planning for magazines and newspaper wholesalers. *European Journal of Operations Research*, 90 pp. 1-12.

Dillmann, R. (1999): Tourenoptimierung. In: *dnv – Der neue Vertrieb* 6/7 1999, 22-27. Presse Fachverlag, Hamburg

Dillmann, R. (2000): Früher liefern – öfter fahren. In: *dnv – Der neue Vertrieb* 11/2000, 18-21. Presse Fachverlag, Hamburg.

Dillmann, R. (2002): Strategic Vehicle Routing Problems in Practice – A Pure Software Problem or a Problem Requiring Scientific Advice. In: Klose, A. / Speranza, M. G. / van Wassenhove, L: N. (eds.): Quantitative Approaches to Distribution Logistics and Supply Chain Management. Springer Verlag, Berlin.

Finkenzeller, K. (2003): RFID Handbook. Fundamentals and Applications in Contactless Smart Cards and Identifikation. John Wiley and Sons Ltd.

Hofmann-Wellenhof, B. / Lichtenegger, H. / Collins, J. (2001): Global Positioning System. Theory and Practice. 5. Aufl., Springer Verlag, Wien

Kopfer, H. (2002): Systeme und Prozesse der zeitnahen Logistik. accepted for publication in Proceedings of Braunschweiger Gesamtverkehrsforum 2002

Mansfeld, W. (1998): Satellitenortung und Navigation. Vieweg Verlag, Wiesbaden.

Paret, D. (2002): RFID and Contactless Smart Cards. John Wiley and Sons Ltd

Pflaum, A. (2001): Transpondertechnologie und Supply Chain Management. Edition Logistik, Band 3. Deutscher Verkehrs Verlag, Hamburg

Pressegrosso [2002]: Pressemitteilung des Bundesverbandes Pressegrosso e.V. 06.05.2002. http://www.bvpg.pressegrosso.de/html/download_doc.php?file=Presseerklaerung_060202.pdf. (10.01.2003)

Wolfertz, T. F. H. [2002] Verspätete Anlieferung und Verkauf bei der BILD. Internal Working Paper Dillmann + Co. GmbH, Wuppertal.

Chapter 3

Warehousing

Planning the Incorporation of Logistics Service Providers to Fulfill Precedence- and Time Window-Constrained Transport Requests in a Most Profitable Way

Jörn Schönberger and Herbert Kopfer

Chair of Logistics, Department of Economics and Business, University of Bremen, Germany

Abstract. We investigate an extension of the Pickup and Delivery Problem with Time Windows (PDPTW) in which only a part of the available request portfolio can be fulfilled by own equipment and the remaining requests are served by subcontracted carriers (Logistic Service Providers). The costs for the fulfillment of the request portfolio result in the sum of travel costs for the self-served requests and the charges paid to the Logistic Service Providers. A Memetic Algorithm (MA) is configured to determine solutions of this combined selection, assignment and sequencing problem. Special genetic operators are proposed to identify those requests that should be served by Logistics Service Providers in order to minimize the overall fulfillment costs. By means of suitable benchmark instances, the capability of the algorithmic approach is assessed and the impacts of the selection feature are analyzed.

1 Introduction

In traditional vehicle routing and scheduling problems routes have to be determined in order to fulfill a given set of requests at costs as low as possible using a fleet of vehicles. However, this approach is not adequate anymore for today operations planning of a carrier company.

Former routing and scheduling problems arose from the distribution of goods of a company maintaining an own fleet. Today, the fleet is out-sourced and operates for several companies. Temporal contracts are made or even single requests are served for a certain charge. On the other hand, the Internet-based request acquiring and offering allows an ad-hoc adjustment of the available request stock, especially by incorporating subcontractor for selected requests.

More general planning approaches are required to exploit these new opportunities for carrier companies (Cruijssen (2003)). Beside generating routes, it has to be decided which subset of requests is served with carrier-owned equipment and which requests are unattractive and should be sub-contracted in order to reduce costs.

In a pickup and delivery problem, goods have to be moved between customer specified locations in order to fulfill certain customer demands. Such

B. Fleischmann et al. (eds.), *Distribution Logistics*

a demand is specified within a pickup and delivery request. To reduce the overall travel costs of the considered carrier company, several requests are consolidated into routes. Limited transport resources and time windows that must be incorporated in order to provide an appropriate customer service often hinder the route generation. Requests not served within such a route are fulfilled by another carrier that receives a charge from the original carrier that had acquired the requests.

The partition of the request portfolio into the set of self-served and into the set of subcontractor-served requests has crucial impacts for the overall costs. This article is about the identification of a cost minimal partition of a portfolio of Less-Than-Truckload (LTL) pickup and delivery requests. Section 2 surveys the general possibilities to incorporate the selection feature into a route optimization model. The Profitable Pickup and Delivery Selection Problem (PPDSP), introduced in section 3, describes the general problem of separating self-served requests from those that are served by a subcontractor. In section 4, a genetic algorithm framework is proposed to identify the most profitable portfolio partition and to consolidate simultaneously adequate routes for the self-served requests. Results obtained from numerical experiments are shown in section 5. This article terminates with a conclusion of the main results and provides some hints for future research topics.

2 Model-Based Request Selection

Subject of this section is the compilation of general modeling approaches to establish a model-based selection of the most profitable requests.

The available transport requests are collected in the set $\mathcal{R}$. A *selection* is an ordered pair $\mathcal{S} = (\mathcal{R}^+, \mathcal{R}^-)$ representing the partition of $\mathcal{R}$ into the set $\mathcal{R}^+$ of self-served requests and the set $\mathcal{R}^-$ of the remaining requests, which are typically served by a sub-contracted carrier. It is denoted as *feasible selection*, if the requests in $\mathcal{R}^+$ can be fulfilled by the equipment of the company so that all given constraints are satisfied.

2.1 Determination of Costs and Revenues

Let $C(\mathcal{S})$ represent the costs associated with the execution of routes to serve the requests accepted according to $\mathcal{S}$. The considerable costs $L(\mathcal{S})$ arising from the accepted requests mainly consists of the travel expenses. They are in general depending upon the length of the routes executed by the vehicles within the available fleet. To minimize $L(\mathcal{S})$, the requests in $\mathcal{R}^+$ are consolidated into routes so that existing constraints on the available capacity, visiting times and visiting orders are satisfied. The minimization of $L(\mathcal{S})$ is a routing and scheduling problem. It is referred to as the routing problem.

Requests contained in $\mathcal{R}^-$ are not served by the available fleet. If the rejection of requests is not possible or not recommended then the requests in $\mathcal{R}^-$ are served by an external carrier service that is paid for fulfilling certain transport demands. The paid amounts $F(\mathcal{S})$ are denoted as *carrier service costs* or *charges*. Different problems concerning the determination of a requests selection that separates the self-served requests from those, which are served by carrier services are investigated by Diaby and Ramesh (1995) for a pure distribution, respectively Kopfer and Pankratz (1998), Greb (1998) and Schönberger et al. (2002) for problems with pickup and delivery requests.

If the consolidation of externally served requests into bundles leads to savings due to degressive carrier service charges, a Freight Optimization Problem (FOP) must be solved (e.g. Kopfer (1984)).

The determination of the least costs $C(\mathcal{S}) := L(\mathcal{S}) + F(\mathcal{S})$ for serving a selection $\mathcal{S}$ therefore requires the solving of a combined sequencing and assignment problem (the routing problem) for the accepted requests and the solving of an FOP for the rejected requests (Pankratz (2002)).

The amount of gained revenues $R(\mathcal{S}) := A(\mathcal{S}) + E(\mathcal{S})$ includes the revenues $A(\mathcal{S})$ associated with each self-served request and, in case of a carrier service incorporation, the revenues $E(\mathcal{S})$ of requests served by external carriers. Subcontracting is often not desired and leads therefore to reduced revenues for externally served requests.

The determination of the selection $\mathcal{S}$ remains an open issue. To decide whether a request r belongs to $\mathcal{R}^+$ or $\mathcal{R}^-$, a combined selection (of self-served requests), assignment (of self-served requests to vehicles) and sequencing (of the visiting locations) problem has to be solved. In the following subsections 2.2 - 2.5 general approaches are presented to merge both goals revenue maximization and costs minimization into one mathematical programming decision model. The corresponding literature for LTL pickup and delivery request consolidation is surveyed.

2.2 Cost-Constraint Selection

The realization of routes comprising the most promising requests in $\mathcal{R}^+$ is typically hindered by a budget B_{up} of available amounts of money, time or other consumed resources. As long as the budget is not bailed out, additional requests can be incorporated within the routes. If the budget is exhausted, the routes must be reorganized to fulfill the so far incorporated requests at lower costs.

The goal is then to find a requests selection $\mathcal{S}$ so that the overall achieved benefit is maximized respecting the available budget. In terms of a formal optimization model, the outlined problem can be written as

$$\begin{aligned} & \max R(\mathcal{S}) \\ s.th. \quad & C(\mathcal{S}) \leq B_{up}, \\ & \mathcal{S} \text{ is a feasible selection,} \end{aligned} \tag{1}$$

The problem (1) is a combined request selection and routing problem, constrained by a restricted budget. It represents the problem of selecting the most promising requests if at least one constrained resource is incorporated. This principle of selection is labeled as **cost-constraint selection**. In general, the costs are allowed to be split into several components each restricted by a budget. It must be ensured, that every budget is not exceeded.

Cost-constrained selection is applied to vehicle operations planning if the available transport capacities are scarce and/or tight time windows hinder a sequential request fulfillment. Schönberger and Kopfer (2003) study the impacts of scarce transport capacity in an LTL pickup and delivery scenario.

2.3 Fulfillment Selection

Some vehicle routing problems aim at determining a request selection to satisfy a certain goal, expressed in the least benefit quantity B_{low}. The realizable benefit is typically not known in advance and it cannot be ensured that a certain benefit quantity B_{low} is realizable, especially if the costs for achieving B_{low} are limited by an upper bound. In such a case it is more promising to try to minimize the costs for realizing the predefined goal B_{low}. The problem (2) represents the task to search for a least cost selection $\mathcal{S}$ to satisfy the predetermined goal B_{low}. This selection principle is referred to as **fulfillment selection.**

$$\begin{aligned} & \min C(\mathcal{S}) \\ s.th.\ & R(\mathcal{S}) \geq B_{low}, \\ & \mathcal{S} \text{ is a feasible selection,} \end{aligned} \tag{2}$$

(3)

Since the specified budget hinder the visitation of every available location, the requests are separated into those, which are visited and those that are not visited. Often, additional restrictions must be taken into account in order to ensure that all unvisited locations lie within an acceptable distance to the nearest visited location. This is important in order to ensure a reasonable service for unvisited customers (e.g. in case of routing of mobile health care units in developing countries).

2.4 Profit Maximization

If B_{up} is fixed too low, some promising requests have to be ignored. A determination of the goal B_{low} at a too low level also leads to a refusal of promising requests whereas an overestimated goal enforces the incorporation of unprofitable requests.

One possibility to overcome with these deficiencies is to leave both costs and benefits unbounded and to search for a selection, which maximizes the overall profit contribution, defined as the difference between the collected revenues and necessary costs. The task is then to identify those requests, whose incorporations lead to positive contributions to the profit of the considered carrier company, representing its success.

$$\max R(\mathcal{S}) - C(\mathcal{S}) \quad (4)$$
$$s.th.\ \mathcal{S} \text{ is a feasible selection.}$$

The principle of request selection described just above is called **profit maximization.**

The selection of the most profitable pickup and delivery requests is described for the single vehicle case by Verweij and Aardal (2000), whereas Schönberger et al. (2002) investigates the multi-vehicle case.

Frantzeskakis and Powell (1990) and Kleywegt and Papastavrou (1998) investigate the problem of accepting full truckload pickup and delivery requests.

2.5 Multi-Objective Formulations

If the benefits and costs are incompatible and cannot be merged into a single objective, profit maximization is not possible. In such a case it is often first tried to find an auxiliary measurement for the benefit or the costs, which substitutes at least one of the aspects, so that both are again compatible.

A bi-criteria-formulation is necessary if no adequate substitution is possible. The determination of a selection that fulfills both single criteria, benefit maximization and cost minimization, at reasonable levels is aimed at. These selections are situated on the so-called Pareto frontier or efficient frontier. They occupy the property that an improvement of one goal (e.g. costs) requires degradation of the other goal (e.g. benefits) and vice versa. In terms of a mathematical programming formulation, this situation can be formulated as

$$opt\ Z(\mathcal{S}) = (R(\mathcal{S}), C(\mathcal{S}))$$
$$s.th.\ \mathcal{S} \text{ is a feasible selection.} \quad (5)$$

Solutions of this problem optimize the vector-value objective function Z. This kind of request selection is denoted as **pareto selection.**

Different goals can be combined in bi-criteria request selection models. Promising combinations are: travel duration and travel costs for a travel cost function that increases if travel speed is accelerated, collected requests and travel costs, travel distance and minimal distance to the visited routes.

3 Profitable Pickup And Delivery Selection Problem

The remainder of this article is dedicated to a special LTL pickup and delivery request selection problem. In the following, it is distinguished between carrier-owned vehicles whose routes can be determined and vehicle of subcontracted logistics service providers (LSP). The route of an LSP-vehicle cannot be affected. A freight charge is paid to an LSP for each served request. It is assumed that an LSP is available for each request. A request that is not selected for being served by a carrier-owned vehicle is assigned to an LSP that becomes responsible for its reliable fulfillment. A charge is paid to the LSP. Since all requests are served, the sum $R(\mathcal{S})$ of revenues is fixed independently from the chosen selection $\mathcal{S}$. Therefore, the achieved profit is maximized if the sum of costs (travel costs and freight charges) is minimized.

Combined request selection and route generation problems for less than truckload pickup and delivery requests have received only minor attention so far as seen in the previous section 2.

3.1 Problem Description

Assume a carrier company with m vehicles. The request portfolio $\mathcal{R}$ consists of the available n requests $r_1, \ldots, r_n$. To obtain a maximal profit the most promising requests are consolidated into at most m trips. Each trip is served by exactly one of the available vehicle v with capacity C_v^{max}. Unconsidered requests are given to an LSP, which receives a previously known charge. It is assumed that exactly one LSP is available for each request, so that it is not necessary to select between different LSP-charges.

Every customer request r_i is specified by the triple (PU_i, DL_i, c_i). The pickup activity PU_i takes place at p_i^+ whereas the delivery activity DL_i is demanded at location p_i^-. A time window $[t_{min}, t_{max}]$ is specified for each activity. Load of volume c_i is to be picked up at p_i^+ and to be delivered to p_i^-.

An operation is a triple $\pi := (p, a(p), e(p))$, where p represents the location of a pickup, a delivery, a start or a stop activity. The expression $a(p)$ refers to the determined arrival time of the vehicle at location p and $e(p)$ denotes the leaving time from p. If the vehicle arrives at p before the associated time window has been opened it has to wait until the earliest allowed operation time $t_{min}(\pi)$ for π has passed. The leaving time of π has to precede the latest allowed operation time $t_{max}(\pi)$.

A sequence of operations $\Pi = (\pi_1^\Pi, \ldots, \pi_{n_\Pi}^\Pi)$ is called a route. In the remainder of this article the first component of π_i is denoted by p_i. The route Π includes N_Π requests. The initial starting operation and the final terminating operation are not stored within the route for the sake of simplification.

Let $t_{p_i,p_{i+1}}$ be the travel time between p_i and p_{i+1}. The arrival and the leaving times are calculated recursively: $a(p_1) = e(p_1) := 0$ and $a(p_i) := e(p_{i-1}) + t_{p_{i-1},p_i}$ $(i > 1)$.

The leaving time is achieved by $e(p_i) := \max\{a(p_i), t_{min}(\pi_i)\}$.

The vector $\delta^\Pi = (\delta_1^\Pi, \ldots, \delta_{N_\Pi}^\Pi)$ describes the volumes that are collected along the route Π. For a pickup operation at p_i it is $\delta_i \geq 0$ and for the associated delivery operation at p_j we define $\delta_j := -\delta_i$. The capacity usage along Π is determined recursively: $\omega_1^\Pi := 0$ and $\omega_i^\Pi := \omega_{i-1}^\Pi + \delta_{i-1}$ $(i > 1)$.

The route Π is called pd-path for vehicle v if it holds the following restrictions (cf. Savelsbergh and Sol (1995)). Either both operations of request r or none of them are contained in Π (*pairing*), a pickup operation precedes its associated delivery operation (*precedence*), the maximal load is not exceeded for all i: $\omega_i^\Pi \leq C_v^{max}$ (*capacity*) and the leaving time for operation π_i lies in the specified time window: $t_{min}(p_i) \leq e(p_i) \leq t_{max}(p_i)$ (*time window*).

The set Π^* contains the requests incorporated into the route Π.

A pd-schedule Ω is a set of pd-paths $\Pi_1, \ldots, \Pi_m$ so that each customer request is assigned to at most one of these paths.

Filling $\mathcal{R}^+$ with the requests contained in one of the pd-paths $\Pi_1, \ldots, \Pi_m$ and $\mathcal{R}^-$ with the remaining requests, $\mathcal{S} := (\mathcal{R}^+, \mathcal{R}^-)$ is a request selection. The costs associated to a pd-path only depend upon the traveled distance. For each driven distance unit, one monetary unit is spent. The expression $C(\Pi)$ represents the costs for executing the pd-path Π and $F(\mathcal{R}^-)$ denotes the freight charges paid for the sub-contracted requests that forms the set $\mathcal{R}^-$.

The PPDSP can now be formulated as the mathematical optimization problem

$$\min F(\mathcal{R} \setminus \bigcup_{i=1}^{m} \Pi_i^*) + \sum_{i=1}^{M} C(\Pi_i), \tag{6}$$

$$s.th.\ (\Pi_1, \ldots, \Pi_m) \text{ is a pd-schedule.} \tag{7}$$

3.2 Generation of Benchmark Instances

A set of benchmark instances for the PPDSP is generated adopting an idea found in Nanry and Barnes (2000). The main concept is to derive instances for the Pickup and Delivery Problem with Time Windows (PDPTW) from optimal or near optimal solutions of the famous Solomon instances for the Vehicle Routing Problem with Time Windows (VRPTW) (cf. Solomon (1987)). Additionally, an adequate freight charge is assigned to each generated request that has to be paid if this request is completed by an LSP.

Customer locations are paired randomly within the routes of the considered solution to obtain pickup and delivery requests. The first visited location becomes the pickup location whereas the remaining one becomes the delivery place (Fig. 1). The demand at the selected pickup location becomes the volume to be moved between the pickup and the delivery location.

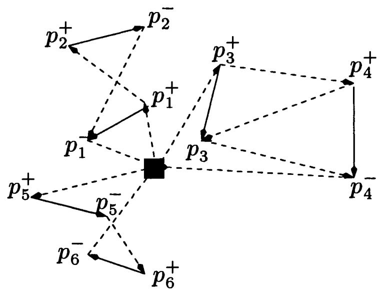


Fig. 1. Generation of a benchmark instance: Derivation of six pickup and delivery requests (solid arcs) from three tours (dashed arcs).

The freight charge q_{r_i} that represents the price for the incorporation of an LSP to complete request r_i is determined with regard to the demanded distance $d(p_i^+, p_i^-)$ between the pickup and the delivery location of this request. Consider the available solution of the Solomon instance. Let v_* be the index of the vehicle serving r, D_{v_*} be the sum of distances between the pickup and the delivery location of the requests assigned to v_* and let L_{v_*} be the driven distance of v_*. For each request r served by v_* the tariff coefficient $\bar{m}$ is set to $\bar{m} := L_{v_*}/D_{v_*}$ and the freight charge is defined as $q_{r_i} := \bar{m} \cdot d(p_i^+, p_i^-)$. It is assumed that the available transport capacities are not scarce. In this case a competition among requests does not take place. Furthermore, the travel costs are only affected marginally by the moved quantities. Greb (1998) proposes a freight tariff that can be applied if the capacities are scarce.

The used solution of the PDPTW-instance is evaluated as a PPDSP solution calculating the overall profit contribution. The achieved value serves as a reference.

Benchmark instances are derived from 18 Solomon VRPTW instances and their solutions. Problems with tight time windows and scattered (R1), semi-scattered (RC1) and clustered customer locations (C1) are considered as well as problems in which the associated time windows are more relaxed (R2, RC2 and C2). From each class, three problems are selected. The pairing is seeded by $\sigma \in \mathcal{G} = \{0, 1, 2\}$.

With the probability α, the LSP-tariff q_r of a certain request r is modified. The updated q_r is given by $q_r := (1 - \beta) \cdot q_r$. If $\beta > 0$, then a discounted tariff is applied otherwise a surcharge has to be paid. Instances $I(\phi, \sigma, \alpha, \beta)$ are generated for the probabilities $\alpha = 0.5, 0.75, 1$ and the discount/surcharge values $\beta = -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75$. Overall, $18 \cdot 3 \cdot 3 \cdot 7 = 1134$ different instances are generated.

The capacity of the incorporated vehicles is set to the sufficiently large value of 300 capacity units.

4 Genetic Search Approach

Evolutionary algorithms and especially Genetic Algorithms have been rarely investigated in the context of solving pickup and delivery type problems. This is caused by a missing problem representation and appropriate operators that allow the simultaneous coding of the assignment, the sequencing and, for PPDSP-type problems, the selection decisions.

Pankratz (2002) proposes a Grouping Genetic Algorithm (Falkenauer (1999)) in which only the assignment of requests to vehicles is left for evolutionary improvement. The routes are constructed then by specialized insertion heuristics.

The applicability of Genetic Algorithms to highly constrained complex optimization problems is often hindered by infeasibility problems. The evolved individuals violate one or more of the problem inherent constraints. To overcome this problem, the combination of GAs with powerful special heuristics to produce feasible solution instances is proposed (Radcliffe and Surry (1994)). The so-called Memetic Algorithms (MAs) often outperform the pure genetic search but, above all, they enable the application of genetic search because they ensure the feasibility of the maintained and evolved individuals.

The so far observed success of MAs motivates the configuration of an MA to tackle the PPDSP. Tentative routes that comply with the pairing and the precedence conditions are obtained from the evolutionary process. They are transformed into the pd-paths removing constraint violating requests. These requests are assigned to an LSP.

4.1 Problem Representation

Encoding A pd-schedule Ω consists of m pd-paths $\Pi_1, \ldots, \Pi_m$ and a collection of rejected requests. Let T_{Π_i} denote the sequence of operations executed by vehicle i. The dummy route T_+ of length N_+ includes the sub-contracted requests. To achieve a string representation $G(\Omega)$ of Ω the sequences are subsequently written in a vector, followed by the dummy route, the lengths of the sequences and of the dummy route and the vector $(D_1, \ldots, D_m)$ of selected termination depots. Vehicle i finishes its operations at the depot number D_i. The representation of Ω is then given by

$$G(\Omega) := (\underbrace{T_{\Pi_1}, \ldots, T_{\Pi_m}, T_+}_{\text{route segment}}, \underbrace{2N_{\Pi_1}, \ldots, 2N_{\Pi_m}, N_+}_{\text{number segment}}, \underbrace{D_1, \ldots, D_m}_{\text{depot segment}}). \quad (8)$$

If requests are shifted between pd-paths or between a pd-path and the dummy route, the length of $G(\Omega)$ does not vary. The length of the route segment is $2n$; the number segment has the length m+1 whereas the depot segment consists of m components. Altogether, the pd-schedule is coded in a string of length $2(n+m)+1$.

Decoding To decode a string representation into a pd-schedule it is proceeded as follows. Tentative routes are derived from the string. Such a route holds the pairing and the precedence constraints but not necessarily the capacity or time window conditions. Initially, the number N_1 of requests assigned to vehicle one is derived from the string representation. The first N_1 pickup operations and the corresponding delivery locations in the route segment form the tentative route of vehicle number one. The tentative route of the second vehicle consists of the next N_2 pickup and its corresponding delivery operations executed in the order determined in the string representation. The tentative routes for the remaining vehicles are simultaneously generated.

In the second step, the tentative routes are converted into pd-paths. First, a modified 2-opt improvement procedure is applied in order to reduce the travel distance and travel time and hence to decrease the number of too-late-arrivals. Afterwards, all requests that cause a capacity or time window constraint violation are successively shifted from the tentative routes into the dummy route and remain unconsidered in the current pd-schedule. The obtained routes now fulfill also the capacity and time window constraints, thus they are pd-paths. The order in which the requests are checked is externally determined. Finally, the 2-opt procedure is re-applied in order to achieve additional travel distance savings. The second step is referred to as repair step in the remainder of this article.

4.2 Operators

Initial Population To generate an initial population of pd-schedules, the parallel path construction heuristic described in detail in Schönberger and Kopfer (2003) is applied. This procedure is parameterized by a permutation of the available requests and by a permutation of the vehicles. It is applied with different request and vehicle permutations to obtain a diversified population.

Crossover The crossover operator derives a new pd-schedule (offspring), coded in a string representation, from two parental pd-schedule representations $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$. First experiments with a syntactical crossover and a syntactical mutation operator (cf. Schönberger et al.(2002)) have not led to convincing results. Therefore, problem specific knowledge is used for generating new pd-schedules.

Initially, the string representations of the parental pd-schedules are split into the routes for the vehicles and into the dummy route. For each individual, $m+1$ routes are then available (some of them are contingently empty).

Offspring routes are generated successively for each vehicle. The parental routes of the vehicle i are denoted as r_i^1 and r_i^2 with l_i^1 and l_i^2 included locations. Let δ_i^{12} be the number of stops that are included in both routes. The offspring route r_i^{off} of length $l_i^{off} := l_i^1 + l_i^2 - \delta_i^{12}$ is initialized. To fill this route it is proceeded as followed. A binary string $\boldsymbol{b} = (b_1, \ldots, b_{l_i^{off}})$ is generated at random. A '0' in the l-th position indicates, that the l-th stop is taken from r_i^1, in case of $b_l = 1$ the next stop included in the offspring route is taken from r_i^2. The probability to select '0' for an arbitrary position in $\boldsymbol{b}$ is set to $\frac{l_i^1 - \delta_i^{12}/2}{l_i^{off}}$, that is the relative length of r_i^1 with regard to r_i^{off}. Starting from $l = 0$ each position of r_i^{off} is successively filled, distinguishing three cases:

1. There are unconsidered stops in both parental routes; Set the l^{th} stop in r_i^{off} to the first so far unconsidered stop in r_i^1 ($b_l = 0$) or r_i^2 ($b_l = 1$). These stops are labeled as considered in both parental routes. It is continued with the next stop ($l := l + 1$).
2. If r_i^1 contains no more unconsidered stops, the remaining stops in the offspring route are filled with the so far unconsidered stops from r_i^2.
3. If r_i^2 contains no more unconsidered stops, the remaining stops in the offspring route are filled with the so far unconsidered stops from r_i^1.

This crossover operator produces a new path that fulfills the pairing and the precedence constraint. Additionally, it does not destroy sequences appearing in both parental routes. A precedence relation of two locations that is included only in one parental route, maybe r_i^1, survives with the probability $\frac{l_i^1 - \delta_i^{12}/2}{l_i^{off}}$.

Each request included into an offspring route is labeled as used and cannot be considered for another subsequently generated offspring route. This ensures that no request is served more than once.

After determining all offspring routes (including the route of the so far unassigned requests) they are stored into the chromosome representation and the number segment is updated. Finally the terminating points of the routes are merged applying a uniform crossover operator.

Mutation With a certain probability each offspring is affected by mutation. Three slight changes are performed.

1. The termination point of an arbitrarily selected route is replaced at random.
2. Within a randomly selected route (including the dummy route) an arbitrarily selected location is re-positioned at random. The precedence feasibility is not violated.
3. A request is moved from the dummy route to the route of an arbitrarily selected vehicle. Therefore, a randomly selected request is deleted from the dummy route and inserted in the route of the determined vehicle at random, so that the precedence feasibility is preserved.

Feature number 3 is the counterpart of the repair step. The latter one removes requests from the current schedule whereas the former feature injects additional requests into the current schedule.

Determination of the Fitness After the application of the repair step all individuals are feasible with respect to the pairing, the precedence, the capacity and the time windows constraints and no customer requests is assigned to more than one vehicle. A suitable fitness value is then obtained by the objective value determined according to (6).

Selection A $\mu+\lambda$ scheme (Baeck (2000)) is used to derive a new population of individuals. In a first step a temporal population is filled with the offspring individuals generated from the parental individuals from the original population. Then the individuals from the original and from the temporal population are sorted in one list by decreasing fitness values. The best individuals within the list form the new population. It substitutes the original population. The population size remains unchanged throughout the generations.

5 Computational Experiments

Several numerical experiments have been performed in order to asses the capability of the proposed genetic search framework and in order to analyze the impacts of varying LSP charges.

5.1 Algorithmic Setup

The Memetic Algorithm (MA) evolves a population of 100 individuals. The initial population is seeded by randomly generated request permutations. While generating the request sequences it is ensured, that requests with surcharge tariffs q_i have a higher probability to be selected for the first components of the sequences (biased request permutations). The parameterization of the construction heuristic with these permutations leads to a diversified initial population. The available fleet is homogeneous, so that it is not necessary to determine different vehicle permutations. The MA generates a sequence of 200 populations each containing 100 individuals. Computational experiments have shown that a crossover probability of 1.0 and a mutation frequency of 0.5 produce the averagely best results. The repair step applied to each constraint violating offspring is parameterized by a biased request permutation, so that requests for which LSP incorporation is most expensive, are checked first.

5.2 Results

Numerical experiments are performed for each instance of the benchmark field. Since the MA is a randomized procedure average results are taken from three independent runs applied to each of the 1134 instances. Therefore, overall 3402 instances are evaluated.

Table 1 shows the averagely obtained results for the six problem classes if the freight charge is neither discounted nor enlarged ($\alpha = 0$). The first line represents the results compared to the reference objective values. Only for problems in the R2 class, the averagely observed results are lying above the reference values. For all other problem classes, the MA is able to identify reduced cost solutions by incorporating an LSP. The largest improvements are observed for C1 problems in which a cost reduction of 16% is realized. The second line represents the percentages of requests that are fulfilled by LSPs. This percentage increases significantly if the spreading of the pickup and delivery locations decreases. In the third row, the percentage of customer locations at which the serving vehicle can execute the corresponding pickup or delivery operation without waiting time for the opening of the corresponding time window is shown. This value tends to decrease if the spreading of the customer locations is reduced. For problems with completely scattered customer locations (R1 and R2) and for problems with clustered locations, the number of no-wait-operations reduces if the time windows are relaxed. This phenomenon is not observed for problems with semi-scattered locations.

Table 1. Results obtained for different problem classes without discounts or surcharges ($\alpha = 0$, $\beta = 0$).

group	R1	R2	RC1	RC2	C1	C2
relative costs	0.98	1.03	0.98	0.91	0.84	1.00
LSP-served	0.19	0.15	0.21	0.27	0.33	0.39
no-wait-service	0.70	0.68	0.65	0.73	0.63	0.50

Table 2 shows the results obtained for experiments with diversified LSP-charges. In all tables, the divergences from the case without discounted or surcharged values ($\beta = 0$) are shown. The tabular on top shows the variation of the sum of costs for a medium diversification ($\alpha = 0.5$), the middle tabular represents the results from the $\alpha = 0.75$ experiment whereas the last tabular contains the results observed for the complete variation experiment ($\alpha = 1.0$).

Two main observations can be stated. If the surcharge is increased then the overall costs also increase and if the LSP-charge discount is enlarged then the savings also increase. Secondly, if the frequency of surcharge-requests is increased (increase of α) then the additional costs also increase. The achieved savings significantly increase if the frequency of discounted LSP-charges is increased.

The largest savings are realized for problems with relaxed time windows (R2, RC2 and C2). Their savings are significantly larger than the savings observed for the problems with tight time windows. This is mainly caused by a reduced number of self-served requests and therefore by an intensified exploitation of the discounted LSP-charges (cf. Table 3).

If the frequency of expensive LSP-charges is enlarged then additional costs are observed. If the extra charge is too large then the MA recognizes that the corresponding requests can be served by own equipment in a cheaper fashion. However, the savings by not using the LSP is so large that it is not necessary to insert these requests in the most profitable way in the existing routes. For this reason, significant additional costs are observed for very large surcharges.

Table 2. Overall costs for the request fulfillment

	medium variation ($\alpha = 0.5$)						
	surcharge ($-\beta$)				discount (β)		
	0.75	0.50	0.25	0.00	0.25	0.50	0.75
avg. R1	0.08	0.04	0.04	0.00	-0.05	-0.14	-0.27
avg. R2	0.13	0.12	0.03	0.00	-0.07	-0.20	-0.35
avg. RC1	0.12	0.08	0.06	0.00	0.01	-0.09	-0.22
avg. RC2	0.11	0.10	0.01	0.00	-0.10	-0.22	-0.35
avg. C1	0.09	0.05	0.01	0.00	-0.06	-0.10	-0.24
avg. C2	-0.01	0.07	0.06	0.00	-0.04	-0.14	-0.27
	medium variation ($\alpha = 0.75$)						
	surcharge ($-\beta$)				discount (β)		
	0.75	0.50	0.25	0.00	0.25	0.50	0.75
avg. R1	0.09	0.08	0.05	0.00	-0.08	-0.21	-0.36
avg. R2	0.14	0.12	0.10	0.00	-0.08	-0.25	-0.47
avg. RC1	0.12	0.10	0.06	0.00	-0.03	-0.13	-0.29
avg. RC2	0.12	0.08	0.03	0.00	-0.13	-0.29	-0.46
avg. C1	0.13	0.06	0.01	0.00	-0.10	-0.19	-0.32
avg. C2	0.08	0.03	-0.01	0.00	-0.06	-0.22	-0.39
	complete variation ($\alpha = 1.0$)						
	surcharge ($-\beta$)				discount (β)		
	0.75	0.50	0.25	0.00	0.25	0.50	0.75
avg. R1	0.14	0.09	0.06	0.00	-0.08	-0.24	-0.43
avg. R2	0.23	0.15	0.09	0.00	-0.12	-0.35	-0.61
avg. RC1	0.16	0.08	0.08	0.00	-0.04	-0.14	-0.32
avg. RC2	0.14	0.15	0.05	0.00	-0.16	-0.38	-0.63
avg. C1	0.15	0.08	0.03	0.00	-0.08	-0.19	-0.35
avg. C2	0.09	0.07	-0.00	0.00	-0.09	-0.26	-0.50

Table 3. Self-served requests ($\alpha = 1.0$ case).

complete variation ($\alpha = 1.0$)							
	surcharge				discount		
	0.75	0.50	0.25	0.00	0.25	0.50	0.75
avg. R1	0.15	0.14	0.06	0.00	-0.22	-0.49	-0.64
avg. R2	0.25	0.26	0.12	0.00	-0.44	-0.81	-0.85
avg. RC1	0.09	0.10	0.06	0.00	-0.08	-0.23	-0.54
avg. RC2	0.67	0.41	0.38	0.00	-0.59	-0.77	-0.82
avg. C1	0.07	0.06	0.02	0.00	-0.14	-0.29	-0.53
avg. C2	0.32	0.24	0.34	0.00	-0.39	-0.62	-0.78

6 Conclusions and Future Works

We addressed the partition of a portfolio of pickup and delivery requests into a set of self-served and into a set of externalized requests. Three one-modal modeling approaches are found: selection caused by limited budgets, least input fulfillment and profit maximization. The latter one is appropriate if no given budget or goal is specified.

The Profitable Pickup and Delivery Selection Problem represents an adequate problem formulation of a combined route generation and request selection striving for the minimization of the request fulfillment costs. The proposed Memetic Algorithm achieves reasonable and comprehensible results.

An interesting problem occurs if online-instances of the PPDSP must be handled. These problems are subject of our current research activities.

References

Bäck, T. / Fogel, D. B. / Michalewicz, Z. (2000): *Evolutionary Computation 1 – Basic Algorithms and Operators.* Institute of Physics Publishing, Bristol and Philadelphia.

Cruijssen, F. (2003): A survey on European inter-organizational data sharing implementations in transportation. Internal report, CentER Applied Research, Tilburg University.

Diaby, M. / Ramesh, R. (1995): The distribution problem with carrier service: A dual based penalty approach. *ORSA Journal on Computing*, 7:24–35.

Falkenauer, E. (1999): *Genetic Algorithms and Grouping Problems.* Wiley, Chichester.

Frantzeskakis, L. F. / Powell, W. B. (1990): A successive linear approximation procedure for stochastic, dynamic vehicle allocation problems. *Transportation Science*, 24:40–57.

Greb, T. (1998): *Interaktive Tourenplanung mit Tabu-Search.* Dissertation University of Bremen.

Kleywegt, A. J. / Papastavrou, J. D. (1998): Acceptance and Dispatching Policies for a Distribution Problem. *Transportation Science*, 32:127–141.

Kopfer, H. (1984): *Lösung des Frachtoptimierungsproblems im gewerblichen Güterfernverkehr*. Dissertation Universität Bremen.

Kopfer, H. / Pankratz, G. (1998): Das Groupage-Problem kooperierender Verkehrsträger. In: Kall, P. / Lüthi, H.-J. (eds.), *Proceedings of Operations Research 1998*, pp. 453–462.

Nanry, B. / Barnes, J. W. (2000): Solving the pickup and delivery problem with time windows using reactive tabu search. *Transportation Research Part B*, 34:107–121.

Pankratz, G. (2002): *Dynamische speditionelle Transportdisposition unter besonderer Berücksichtigung der Fremdvergabe*. DUV, Gabler Ed. Wissenschaft, Wiesbaden.

Radcliffe, N. J. / Surry, P.D. (1994): Formal Memetic Algorithms. In: Fogarty, T. C. (ed.), *Evolutionary Computing*, vol. 865 of *Lecture Notes in Computer Science*, pp. 1–16. Springer-Verlag, Berlin, Heidelberg, New York.

Savelsbergh, M. W. P. / Sol, M. (1995): The general pickup and delivery problem. *Transportation Science*, 29:17–29.

Schönberger, J. / Kopfer, H. (2003): A memetic algorithm for a pickup and delivery selection problem. In: Mohammadian, M. (ed.), *Proceedings of International Conference on Computational Intelligence for Modeling Control and Automation*, published on CD.

Schönberger, J. / Kopfer, H. / Mattfeld, D. (2002): A combined approach to solve the pickup and delivery selection problem. In: Leopold-Wildburger, U. / Rendl, F. / Wäscher, G. (eds.), *Proceedings of Operations Research 2002*, pp. 150-155.

Solomon, M. M. (1987): Algorithms for the vehicle routing and scheduling problem with time window constraints. *Operations Research*, 35:254–265.

Verweij, B / Aardal, K. (2000): The Merchant Subtour Problem. Research Report UU-CS-2000-35, Utrecht University.

Williams, H. P. (1999): *Model Building in Mathematical Programming*. Wiley, Chichester, fourth ed.

Vehicle On-line Dispatching Rules in Practice

Tuan Le-Anh and René De Koster

Erasmus University Rotterdam, Rotterdam School of Management/ Faculteit Bedrijfskunde, Rotterdam, The Netherlands

Abstract. Vehicle-Based internal transport systems are widely used in many industrial facilities such as warehouses, distribution centers and production facilities. In most vehicle-based internal transport systems, on-line dispatching rules are used to control vehicle movements. Single-attribute dispatching rules, which dispatch vehicles based on one parameter such as the load waiting time, are used popularly in practice. Examples of these rules are modified-first-come-first-serve and nearest-workstation-first rules. Hybrid dispatching rules are more complicated and dispatch vehicles using several parameters and types of vehicle assignments. The major advantage of single-attribute rules is that they are simple and easy to implement in practice, but their performance can be improved. In this research, we propose new dispatching rules and compare their performance with several single-attribute and multi-attribute dispatching rules in literature. Experimental results show that our new dispatching rules are robust and perform very well in general.

Keywords. On-line dispatching, vehicle-based internal transport system, centralized control, performance.

1 Introduction

In nearly every industrial facility, material handling systems are being used to transport products (or intermediate products) between internal locations. These facilities can be manufacturing plants, warehouses, distribution centers or transshipment terminals. In these facilities, material handling systems are responsible for transportation of loads between storage locations, dock doors or machine centers. Internal transport is crucial for achieving short lead times, just-in-time production, and reliable deliveries. Conveyors and vehicles are the most commonly used internal transport equipments. Conveyors are appropriate when the goods are conveyable, distances are short, transport batch sizes are small combined with rather high volumes. Conveyors are static and have little flexibility. Vehicles such as forklift trucks, automated guided vehicles (AGVs) have higher flexibility in routing and in the material that can be transported.

This research concentrates on the internal transport environments in which vehicles are main means of transport - Vehicle Based Internal Transport (VBIT) systems. Internal transport environments are usually stochastic and it is nearly impossible to schedule vehicles far in advance. In this case, an on-line strategy is a solution to solve this problem. On-line dispatching systems using on-line

B. Fleischmann et al. (eds.), *Distribution Logistics*

dispatching rules are very popular in vehicle-based internal transport systems and improving their efficiency is an important task.

In this paper, we experiment with different types of on-line dispatching rules to control VBIT systems in two real case studies. These dispatching rules include single-, multi-attribute and hybrid dispatching rules. Details about dispatching rules can be found in the experimental setup section. We also compare the performance of dispatching rules and rank them according to their performance. Performance criteria are minimizing the average load waiting time while the maximum load waiting time is preferably as small as possible.

This paper is organized as follows: the literature on on-line dispatching is studied first, and then cases are described. Next, the experimental set-up is described in detail followed by a performance evaluation for dispatching rules. Finally, some conclusions are given in the last section.

2 Literature Survey

According to the literature, on-line dispatching systems can be divided into two main categories: decentralized and centralized control systems. The decentralized system dispatches vehicles based on only local information available at the decision moment. The centralized control system dispatches vehicles using local information and also uses information available at the central controller. Although recently, some research has been devoted to local agent-based vehicle control (Lindeijer, 2003), in practice, due to their efficiency, centralized dispatching systems are more popular. Depending on the ways in which transportation requests are assigned, the dispatching rules can be divided into two categories (Egbelu and Tanchoco, 1984): workstation-initiated dispatching rules (jobs at workstation have the priority to claim vehicles) and vehicle-initiated dispatching rules (vehicles have the priority to claim jobs). Some common dispatching rules in literature are the shortest-travel-distance-first (STTF), the first-come-first-served (FCFS), modified-first-come-first-served (MODFCFS), the maximum-outgoing-queue-size (MOQS) and the minimum-remaining-outgoing-queue-space (MROQS) rules (see Egbelu and Tanchoco, 1984; Egbelu, 1987; Srinivasan et al., 1994). These dispatching rules are rather simple and dispatch vehicles based on only one parameter such as travel distance.

Klein and Kim (1996) propose several multi-attribute dispatching rules. The dispatching rules presented in their paper are based on the multi-criteria decision making approach and include: the simple additive weighting, the Yager's multi-attribute decision making, the modified additive weighting, and the Max-Max approaches. Factors, which are considered in their decision function, include waiting time, queue length and vehicle travel time. They compare performance of dispatching rules (single- and multi-attribute) using simulation and conclude that the multi-attribute dispatching rules are superior to many single-attribute dispatching rules for most performance criteria, such as average and maximum load waiting times. However, the implementation of these multi-criteria rules is quite complicated and as such they have only been tested in small models. Furthermore, in their research, they did not consider reassigning vehicles (moving-to-park or moving-to-pickup), which certainly lead to improvement of the system performance. Van der Meer (2000) has tested reassigning moving vehicles to loads for MODFCFS and NVF (nearest vehicle first) dispatching rules for simple layouts. Fleischmann et al. (2003) propose three dynamic scheduling procedures using assignment rules, an assignment algorithm and an insertion algorithm to dispatch vehicles dynamically. The assignment algorithm (AS), which schedules all open orders (orders already arrived but not started yet) during a planning horizon (the panning horizon length is the latest completion time of the orders in execution), is the best scheduling method. The assignment algorithm is invoked when a new order arrives. Orders that have time windows later than the planning horizon are postponed. Several types of costs are considered in the assignment problem such as costs for loaded and empty moves, costs for waiting and delay. This assignment algorithm is suitable for systems where information about future pick up is known. Bozer and Yen (1996) propose two dispatching rules, which consider reassignment of moving vehicles to loads. These are modified-shortest-travel-time-first (MOD STTF) and bidding-based device dispatching (B^2D^2). The B^2D^2 is a better rule but is complicated to implement in practice. These dispatching rules aim at reducing the empty vehicle travel time, but can lead to a high value of the maximum load waiting time, since these rules pay no attention to load waiting times.

This research was inspired by the previous research from Van der Meer (2000). Van der Meer (2000) shows that distance-based dispatching rules such as nearest-workstation-first (NWF) perform very well in practice to reduce the average load waiting time. However, we found that these rules also maximize the maximum load waiting time at the same time, and hybrid rules may help to solve this problem. In this paper, we introduce several new dispatching rules. They are simple enough to be implemented in practice and moreover they turn out to be efficient rules in practical environments. Using simulation models of two practical cases, we compare their performance with two commonly used single-attribute rules (MODFCFS and NWF), a multi-attribute rule from Klein and Kim (1996) and a MOD STTF (Bozer and Yen, 1996). We will show that these rules are robust and perform very well in general.

The two case studies are described in the next section.

3 Description of the Implementation Environments

3.1 The European Distribution Center (EDC)

This case concerns the transportation of pallet loads at the European distribution center of a computer hardware and software wholesaler. This wholesaler distributes computer products to different retail stores in Europe and determines how much to purchase and store to be able to comply with the demand of the retailers. Because computer products change quickly over time, it is necessary to keep inventory levels low and the storage times as short as possible. A large part of the incoming products are packed in cartons, stacked per product on pallets. Five forklifts with vehicle-mounted terminals transport the pallets. A central warehouse management system (WMS) keeps track of inventory and the position of stored products. The distribution center can be divided into several areas (see Fig. 1) with a total operating area of 40 by 140 meters. Each weekday, trucks arrive at the *Receiving Lanes* of the distribution center where the pallets (loads) are unloaded. In total there are five Receiving Lanes. If the cartons on the pallets contain returned or broken products they are manually transported to one of the five *Return stations*. The pallets are manually transported to one of twelve *Check-in Area* stations if the content of the cartons is unclear. At each of the previously mentioned stations, the pallets are labeled with a so-called license plate number (bar code). This license plate number contains information about the content of the cartons and the location the pallet should be brought to. At the moment the license plate is placed on the pallet, the pallet is entered into the WMS.

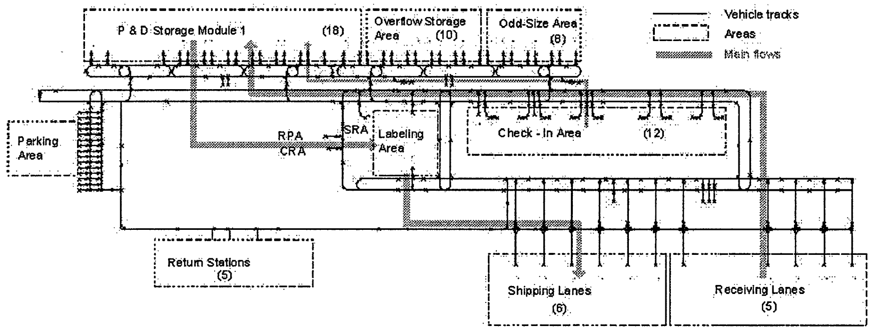


Fig. 1. FLT path layout

If the cartons on the pallet are odd-shaped, or if the pallet is one of many with the same product, it will be transported to the *Odd-Size* or *Overflow Storage Area.* The Odd-Size Storage Area and the Overflow Storage Area have 10 and 8 Pick & Drop (P&D) locations. Otherwise the pallets go to one of the 18 P&D locations of *P&D Storage Module 1.* Within the storage modules, pallets are stored and orders are picked. From Storage Module 1, pallets can be transported to the Repalletization Area (*RPA*), the Shelf Replenishment Area (*SRA*), the Central Return Area (*CRA*), the *Shipping Lanes* and the *Labeling Area.* The Labeling Area has one delivery station and one pick up station. RPA, CRA and SRA have one station each, and there are 6 shipping lanes in total (see Fig. 1). From RPA, pallets move to Storage Module 1 or to CRA. At SRA the cartons of the pallets are placed on a conveyor belt, and will be transported to the shelf area where products are hand picked. Pallets at CRA always move to Storage Module 1. At the Labeling Area, pallets receive customer stickers and packing lists. The Shipping Lanes are the final stations. There, trucks arrive at dock doors to transport products to retail stores.

Simulation Environment

The load release times at their release locations have been measured for a period of six weeks. It appears that the requests for a certain transport depend highly on the time of day and can be modeled properly using Poisson distributions (the distribution has been tested using a χ^2-test). Each type of transport is independently exponentially generated at its own rate. Each day is in turn divided into four periods. Period 1: from the start of the day until the coffee break, period 2: from the coffee break until lunch, period 3: from lunch until the tea break, and period 4: from the tea break until the end of the working day. These periods are introduced to realistically represent the variation in the inter-arrival rates over the day. For example, in period 4 more loads are transported to the shipping lanes than in period 1. Other parameters such as vehicle speed, pickup times come from careful measurements made at the distribution center. The variation in pick-up and drop-off time of loads is very small so we consider these parameters to be deterministic. **Fig. 2** shows the flow of materials inside the European Distribution center. To measure the performance of dispatching rules, the design of the warehouse (Fig. 1) and other relevant specifications of the warehouse and guided vehicles (GVs) have been modeled in the AutoMod simulation software package (version 10.0).

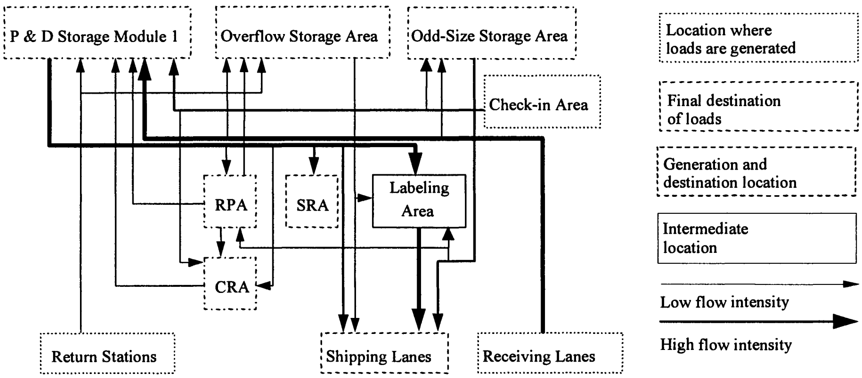


Fig. 2. The material flow between all locations of the EDC

All the parameters are kept the same for each scenario. These parameters include: the material flow, the number of generated loads, time and locations of loads generated in the system, load generation instants, the speed of the vehicles, vehicle capacity, the paths via which the vehicles may travel, the load pick up and set down time, the number of simulated days and the number of working hours per day. Table 1gives a summary of some other values of the simulation model.

Table 1. Parameters

GV speed	2 m/s	Load generation per hour	77
Pick up time of a load	15 s	Current number of vehicles	5
Set down time of a load	15 s	Size of transport area (m×m)	40×140
Vehicle capacity	1 load (pallet)		
Number of working hours per day	7.5 hours	Number of different transport distances (approximately)	800

3.2 The Glass Production Plant

The second case concerns the transportation of pallet loads at a production plant of packaging glass. The glassware is stored after production at the site until the clients (manufacturers that fill the glassware) collect the products for their own use. About 400 different glassware products, varying from jars to bottles, are produced. With three glass melting ovens and nine production lines, nine different glassware products are produced simultaneously, 24 hours per day, 365 days per year. The glassware is carefully stacked on pallets, which are then wrapped in plastic foil and finally moved by three *conveyors* to the 'landing' zone in one of the storage areas, see Fig. 3. There are eight main storage areas (denoted by S1 through S8 in Fig. 3) with a total of 55000 square meters of storage space. Eleven dual-load radio frequency-guided forklifts move two pallets at a time, which arrive at the *conveyors* in pairs and are transported to one of the eight storage areas. The total operating area of the GVs is 315 by 540 meters. The pallets are always moved in pairs. On average between 1200 - 1400 production pallets, 200-250 Value Added Logistics (*VAL*) pallets and 60-70 'extra foil pallets' arrive per day at the landing zone. These inbound pallets are stored by product type in stows of 90-120 pallets.

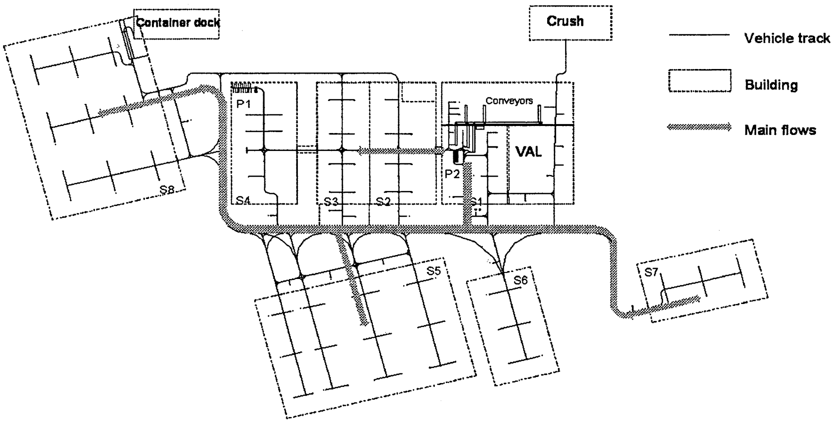


Fig. 3. Forklift path layout connecting all pick up and delivery locations

Simulation Environment

The material flow of the glass production plant is given in **Fig. 4**. On average there are four pallets per day which have to go back to be crushed in the *Crush* area. Within the storage areas, about 200-250 pallets per day have to be moved in batches of 10 pallets to the *VAL* area (except in the weekends) and 200 pallets are reallocated within the storage areas for storage space optimization. Furthermore, on average 1820 outbound pallets have to be moved per day in batches of 28 pallets to 65 trucks which arrive just outside the storage areas between 6.00 am and 10.00 pm, except in the weekends. In 20 % of the cases, the trucks must visit two storage areas to be completely loaded. On average 10 % of all outbound pallets from S8 leave via the *container dock* instead of the main door of S8 since 10 % of the trucks arriving there can only be loaded from the back. Furthermore, there are peak arrivals of trucks during the day, since more trucks arrive in the morning and late afternoon compared to the early afternoon and the evening. The transport vehicles will park at the closest parking place (P1 or P2) when they have no task. The vehicles are free to move anywhere on the paths of the defined operating area (see Fig. 3) and can pass each other if necessary. However, there is room for only one forklift at a time at the pick-up and drop locations of the conveyors, trucks and stows in the storage buildings.

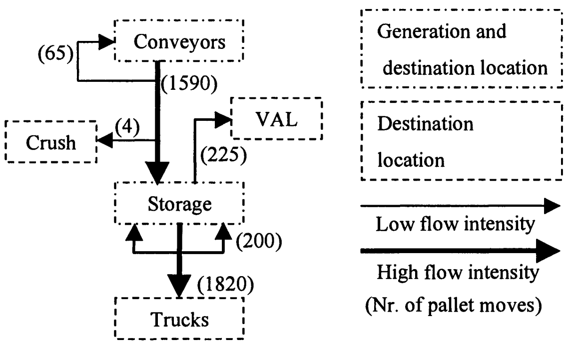


Fig. 4. Average weekday material flow between all locations of the glass production plant

The data on load release times, origins and destinations come directly from the database of the WMS of the company and expert judgments. Other parameters such as vehicle speed, pickup times come from careful measurements made at the production plant. Since pallets are always moved in pairs, the case can be modeled as a uni-load environment with half the number of pallets to be transported. Dual pallets are generated using Uniform inter-arrival times. The uniform character of load generations is due to the uniform release of loads from the production lines. The layout of the production plant (Fig. 3) and other relevant specifications of the environment and forklifts have been modeled in the AutoMod simulation software package (AutoMod version 10). Table 3 gives a summary of some other values of the model.

Table 2. The parameters used for each scenario

Speed of loaded FLTs	on straight paths	2.5 m/s	Vehicle capacity	1 dual-load (pallet)
	in curves	2 m/s	Number of working hours per day	24 hours
Speed of empty FLTs	on straight paths	3.5 m/s	Load generation hour	67(81, 33)
	in curves	3 m/s	Current number of vehicles	11
Pick up time of a load		13 s	Size of transport area (m×m)	315×540
Set down time of a load		14 s	Number of different transport distances (approximately)	2000

4 Experimental Set-up

In this section, we describe characteristics of dispatching rules and experimental environments used in this research.

4.1 Dispatching rules

Egbelu and Tanchoco (1984) divide dispatching rules into vehicle- and load-initiated rules. Vehicle-initiated dispatching rules prioritize the jobs, according to some specific rule. The job that has the highest priority will be selected by a vehicle that becomes idle. Under load-initiated rules, loads have the initiative to claim vehicles using a prioritization rule (vehicles are prioritized for selection). However, a vehicle just finishes a job and has not been claimed by any load, it searches for a load to pick up, using a vehicle-initiated rule. Therefore load-initiated dispatching rules described by Egbelu and Tanchoco (1984) are actually a combination of load- and vehicle-initiated rules in which loads have priority to claim vehicles.

In operation a dispatching rule (load- or vehicle-initiated) is invoked at following events:

- Arrival of a new load,
- A vehicle just finish a job,
- A vehicle is awakened by a load or by another vehicle.

The main difference between the load- and vehicle-initiated dispatching rules is that a load in the system using load-initiated dispatching rules can claim a vehicle for itself. A load in the system using vehicle-initiated dispatching rules cannot claim a vehicle but they can wake a vehicle and then this vehicle will search for a load to pick up. Under vehicle-initiated dispatching rules, when a load enters the system, this load tries to awaken the first vehicle on the vehicle waiting list, which in most cases, is not the best one.

Distance-based dispatching rules such as NWF perform very well to maximize the system throughput when queues' capacities are not restricted (Egbelu and Tanchoco, 1984). Time-based dispatching rules like MODFCFS are good rules in some cases (Srinivasan et al., 1994). Multi-attribute and hybrid dispatching rules, in general, have a better performance in comparison with single-attribute dispatching rules (Klein and Kim, 1996; Bozer and Yen, 1996). Therefore, in this paper we propose three new dispatching rules (one multi-attribute and two hybrid rules) and compare their performance with several good on-line dispatching rules described in literature (NWF, MODFCFS, a multi-attribute rule and MOD STTF).

Characteristics of dispatching rules used in this paper are given in Table 3. For all rules in Table 3, when a vehicle becomes idle (and has not been claimed by a load) and cannot find any load in the system for transportation, this vehicle will park at the closest parking location.

Table 3. Dispatching rules and their characteristics

	Vehicle-initiated	*Workstation-initiated*	*Time priority*	*Reassign-ment*	*Cancella-tion*	*Sources*
Single-attribute dispatching rules						
MODFCFS	✓					Srinivasan et al. (1994)
NWF	✓					Egbelu and Tanchoco (1984)
Multi-attribute dispatching rules						
Multi-att	✓		✓			Klein and Kim (1996)
NWFTP	✓		✓			This paper
Hybrid dispatching rules						
NVF_R	✓	✓		✓		This paper
NWFTP_R	✓	✓	✓	✓		This paper
NVF_RC	✓	✓		✓	✓	Similar to MODSTTF (Bozer and Yen (1996)

Single-attribute dispatching rules

- Modified First-Come-First-Served (MODFCFS)

Under this rule, vehicles have the initiative to claim loads. An idle vehicle looks for loads first at the vehicle current location. If there are several loads here, the vehicle claims the oldest one among them. Otherwise, the vehicle will claim the oldest load in the whole system if this is available. If the vehicle cannot find a load the vehicle stops at its current position (if it is a parking location) or moves to the closest parking location.

- Nearest Workstation First (NWF)

This is a vehicle-initiated dispatching rule. When a vehicle is available, this vehicle claims the closest load (along traveling paths) in a system. If the vehicle cannot find any load, it stays idle (at a parking location) or travels to the closest parking place. The NWF rule is very sensitive to the system layout. Loads in remote areas may never qualify for transportation. This may lead to a high value of the maximum load waiting time in the system.

Multi-attribute dispatching rules

- Nearest Workstation First with Time Priority (NWFTP)

In case waiting times of all loads in the system are smaller than a time threshold θ (load type I), an idle vehicle claims the closest one for transportation. If there are loads with load waiting times larger than θ (load type II), those loads have higher priority for transportation than loads type I. Among type II load, a vehicle selects the nearest one. Loads (type I) are considered only when there are not any loads of type II in the system. The time threshold (θ) is chosen around 3×(average load waiting time when the NWF rule is used). This value was found after several experiments and proved to be good in general.

- Multi-attribute dispatching rules (Multi_att)

Klein and Kim (1996) propose multi-attribute dispatching rules base on several parameters such as the vehicle empty travel distance, the load waiting time and the remaining spaces in queues. In our case studies (section 0) queues are not overflowed under single-attribute dispatching rules, so they are not the bottleneck in these systems. Therefore, mainly vehicle travel distances and load waiting times affect the system performance. Therefore we choose vehicle empty travel distance and load waiting time to be decision attributes. Let dis_{vi} denotes the empty travel distance from the current vehicle (v) location to the pickup location i and $wait_{vi}$ denotes the waiting time of load i. dis_{vi} and $wait_{vi}$ are normalized to DIS_{vi} and $WAIT_{vi}$ using the following expressions:

$$DIS_{vi} = \frac{dis_{vi} - \min_j dis_{vj}}{\max_j dis_{vj} - \min_j dis_{vj}}\ ;\ WAIT_{vi} = \frac{\max_j wait_{vj} - wait_{vi}}{\max_j wait_{vj} - \min_j wait_{vj}}$$

The attributes DIS_{vi} and $WAIT_{vi}$ are used to compute the score function S_{vi}.

$$S_{vi} = w_1 \times DIS_{vi} + w_2 \times WAIT_{vi}\ ;\ w_1 + w_2 = 1$$

w_1, w_2 are weights of the vehicle empty travel distance and the load waiting time respectively.

The score function S_{vi} is then used to select the suitable load for a vehicle. When a vehicle becomes idle, this vehicle searches for a load to pickup as follows:

- If the vehicle finds one or more transportation requests in the system then:
  - Values of the score function for all waiting loads in the system are calculated,
  - A load that has the smallest value of the score function is chosen to be picked up,
- If the vehicle cannot find a job, it goes to the closets parking location and remains idle until being awakened up by a load or by another vehicle.

Van der Meer and De Koster (2000) show that distance-based dispatching rules perform better than time-based dispatching rules, so we give a higher weight for the vehicle empty travel distance attribute. Depending on the specific case, the best attribute weights are different and can be found by experiments. In this case we select the weights of travel distance and waiting time to be 0.8 and 0.2 respectively.

Hybrid dispatching rules

- Nearest Vehicle First with vehicle Re-assignment (NVF_R)

This rule is a workstation-initiated dispatching rule. When a load just enters the system, this load immediately searches for an available idle vehicle or a vehicle that is currently going to a parking location. The vehicle (idle or moving to park), which is closest to the load (distance is calculated along the possible travel paths), will be selected to carry that load. If the load fails to find a vehicle, this load waits at its current location until being claimed by an idle vehicle. When a vehicle becomes idle and is not claimed by a load, the vehicle searches for the closest load in the system.

- Nearest Vehicle First with Time Priority and vehicle Re-assignment (NVFTP_R)

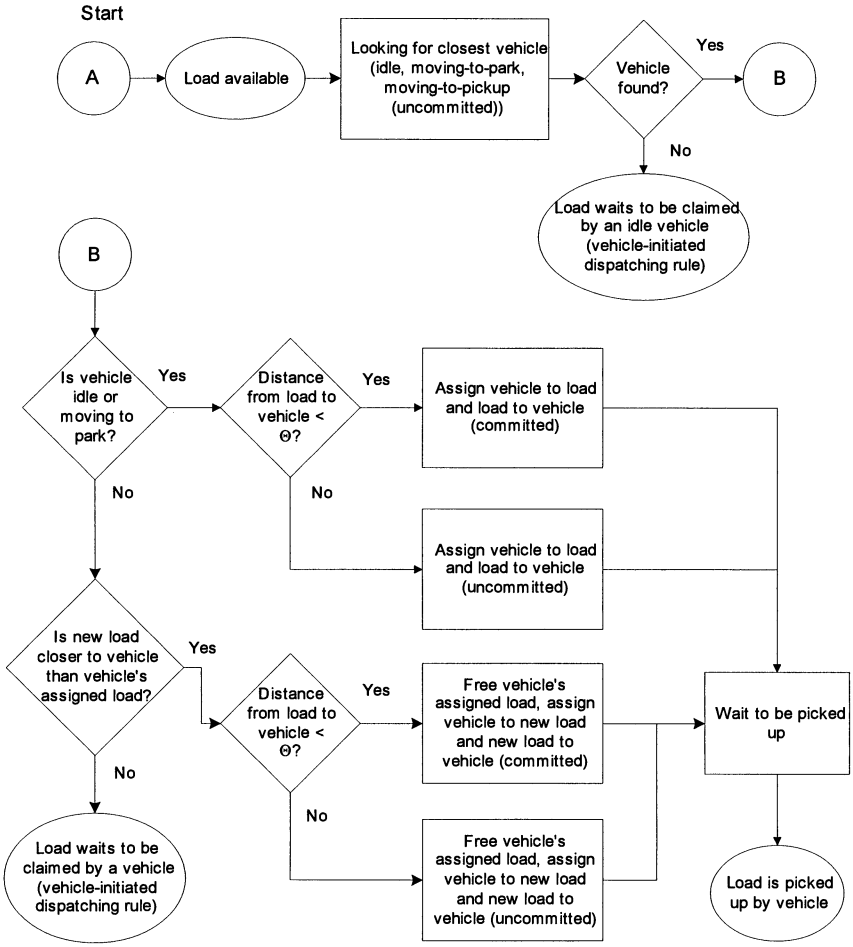
Using this rule, a load that just enters a system claims a vehicle in the same way as a load in a system using the NVF_R rule does. If this load cannot find a vehicle, it remains at its current location and waits until being claimed by an idle vehicle. An idle vehicle claims a load in the same manner as when the NWFTP rule is used. If this vehicle cannot find a load to carry, this vehicle stays idle at it current location (if it is a parking location) or travel to the closest parking place.

- Nearest Vehicle First with vehicle Re-assignment and Cancellation (NVF_RC)

Under NVF_RC, when a load just enters the system, this load claims the nearest vehicle if available (Fig. 5). A load is committed to a vehicle if the vehicle claims the load and the travel distance from the vehicle to the load is smaller than a distance threshold Θ. When a vehicle becomes idle, this vehicle searches for a load as described in Fig. 6.

In Bozer and Yen (1996), in order to select the Θ value, they propose to sort the distances between all pairs of station (in non-decreasing order) and use the 60th percentile as the threshold value. This is a very elaborate procedure. In this research we select Θ value around the average vehicle travel distance when the

NWF rule is used. This value proves to be a good threshold in our two cases. The two threshold values (Θ and θ) are obtained after pilot runs of simulation models using the NWF rule.



*(Vehicle status: **idle**: vehicle stay idle (has no job) at a parking location; **moving-to-park**: a vehicle has no job and is traveling to a parking location; **moving-to-pickup**: a vehicle is traveling to the vehicle's assigned load pickup location; **committed**: means that the vehicle cannot be diverted to another destination, **uncommitted** otherwise.)*

Fig. 5. The impact of load on vehicle behavior in dispatching rules with vehicle reassignment.

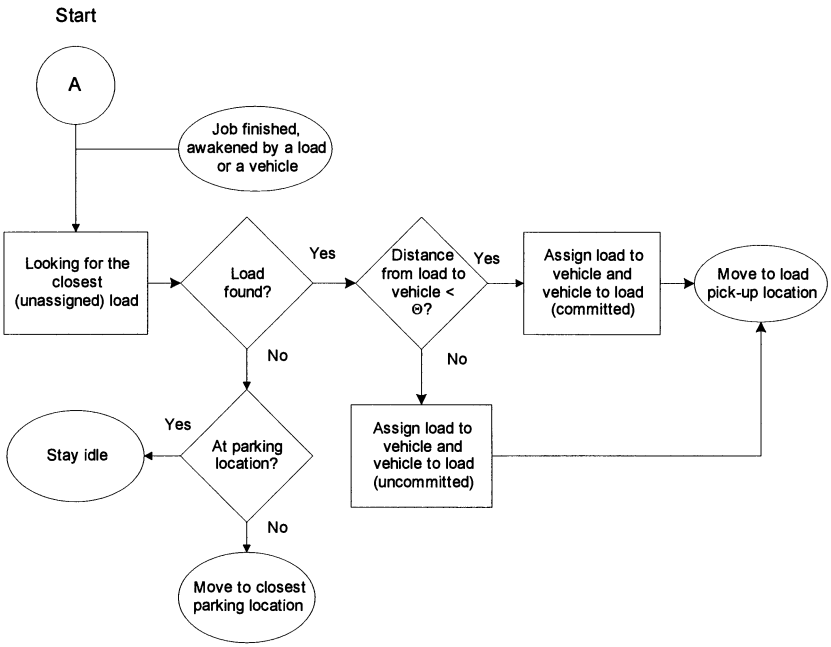


Fig. 6. Vehicle-initiated dispatching

4.2 Experimental environment

These cases have been modeled using AutoMod 10. In simulation models, several assumptions are made:

- The guide-path layout of VBIT systems is given and fixed,
- All vehicles have uni-load capacity
- Vehicles choose the shortest path to pickup and deliver loads
- Vehicles never break down,
- Loads at each input queue are processed on a first come first serve basis,
- The battery-charging problem is neglected,
- Parking locations are given and defined by companies (usually at the location of charging station).

The replication/deletion approach (see Law and Kelton, 2000) was applied for determining mean values of performance indicators. For each combination of experimental factors, a replication of ten runs is used to determine results. Table 4 shows experiment factors used in experiments.

Table 4. Experimental factors

Models	The European Distribution Center	The glass Production Plant
Dispatching rules	MODFCFS, NWF, NWFTP, NWF_R, NWFTP_R, NWF_RC, Multi_att	MODFCFS, NWF, NWFTP, NWF_R, NWFTP_R, NWF_RC, Multi_att
Number of vehicles	4, 5, 6	7, 9, 11

Many performance criteria might be considered for VBIT systems, but in these companies we find that the average waiting time is the most relevant criterion and we use maximum load waiting time only as a supplementary performance criterion. The reason is that for these models the queue spaces are not critical, and the main objective of companies is to transport loads as soon as possible.

In this research, we have done a number of experiments with different scenarios. The experimental factors are described in the Table 4. The results of experiments have been analyzed and rules have been ranked (using Tukey test with an overall confidence level of 95% - $\alpha = 0.05$).

Model validation and verification

The operation of the systems has been validated using 3-D animation together with the responsible manager of the companies. The results have been checked against reality for the current dispatching rules used in these companies. Those dispatching rules are not included in the experiments since their performance are worse than NWF and about the same as MODFCFS.

In the next section, experiments are carried out to evaluate the performance of on-line dispatching rules.

5 Performance Evaluation

In this section, we evaluate the performance of on-lines dispatching rules described in previous section. The main performance criterion is minimizing the average load waiting time while the maximum vehicle waiting time is preferred to be as small as possible. Rules are evaluated with different numbers of vehicles.

5.1 The European Distribution Center (EDC)

5.1.1 Performance for 4 vehicles case

Table 5. Experimental results for the distribution center (4 vehicles)

4 vehicles							
	MODFCFS	NWF	NWFTP	NVF_R	NVFTP_R	NVF_RC	Multi_att
Avewait	1144.62	315.29	386.41	320.03	392.66	319.26	327.96
std	537.58	84.82	148.92	93.81	136.81	111.17	81.27
maxwait	3953.71	4262.68	2222.91	4421.63	2257.62	4372.24	2718.71
VehUtil%	97.86	91.44	91.89	90.63	91.57	> 81.12*	91.59

{Avewait: average load waiting time (s); std: standard deviation; maxwait: maximum load waiting time (s); Vehutil%: vehicle utilization (%) = Percentage of [vehicle travel time (with a job) + vehicle's pick-up &set down time] / total vehicle available time}

** We cannot get the exact number here since it is very difficult to separate "moving-to-pickup" time and empty traveling time in this case.*

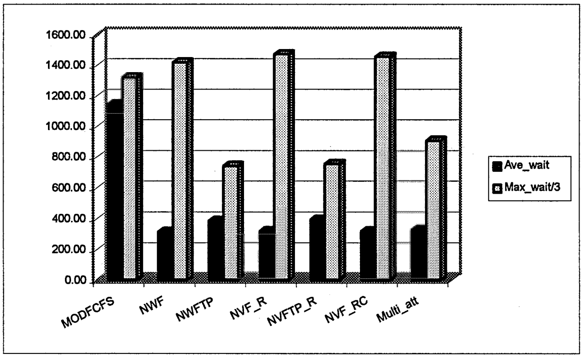


Fig. 7. Performance of dispatching rules when 4 vehicles are used

Table 5 and Fig. 7 show that MODFCFS performs worst in the average waiting time and is not so good in terms of the max waiting time. Considering only the average waiting time, NWF is the best rule in this case. This rule is better than NVF_R and NVF_RC, which should perform well. A reason is that, in this case, the vehicle utilization is very high (about 90%), vehicles are busy most of the time so re-assigning moving vehicles is not always useful. As can be seen in the Fig. 7, the NWFTP rule realizes the small value of the max load waiting time. This is expected, since this rule aims at reducing load waiting times by giving higher priority to loads that have to wait long. In this case, NWFTP, NVFTP_R and Multi_att perform much better than MODFCFS in terms of the max load waiting time. Using the MODFCFS rule leads to vehicle utilization of about 97.6%, which is too high and can lead to an unstable situation.

Table 6. Rank of dispatching rules (Tukey - confidence interval 95%)

Average waiting time				**Max waiting time**		
	Subset				*Subset*	
RULES	1	2		RULES	1	2
NWF	315.29			**NWFTP**	**2222.91**	
NVF_RC	319.26			**NVFTP_R**	**2257.62**	
NVF_R	320.03			**Multi_att**	**2718.71**	
Multi_att	**327.96**			MODFCFS		3953.71
NWFTP	**386.41**			NWF		4262.68
NVFTP_R	**392.66**			NVF_RC		4372.24
MODFCFS		1144.62		NVF_R		4421.63

** Rules in different groups indicate that their mean values are significantly different (with 95% confidence). In ranking tables, rules are marked as good rules (in Bold) if their average and max waiting time and the max waiting time are both in first ranking groups.*

Table 6 shows two groups for average waiting times and two groups for max load waiting times. This table indicates that the average load waiting time is significantly worse when MODFCFS is used, since this rule belongs to the second group. If we compare dispatching rules using the max load waiting time, Multi_att, NWFTP and NVFTP_R perform significantly better than the others. From Table 6, we can pick Multi_att, NWFTP and NVFTP_R as recommended rules for this case.

5.1.2 Performance for 5 vehicles case

Table 7. Experimental results for the distribution center (5 vehicles)

5 vehicles							
	MODFCFS	**NWF**	**NWFTP**	**NVF_R**	**NVFTP_R**	**NVF_RC**	**Multi_att**
Avewait	182.46	134.19	139.51	125.99	130.85	119.81	134.11
std	41.61	14.68	21.88	15.96	19.44	16.62	16.31
maxwait	704.01	1156.99	710.68	1154.44	675.39	1136.92	914.82
VehUtil%	79.00	76.43	76.55	74.77	74.86	> 59.86	76.50

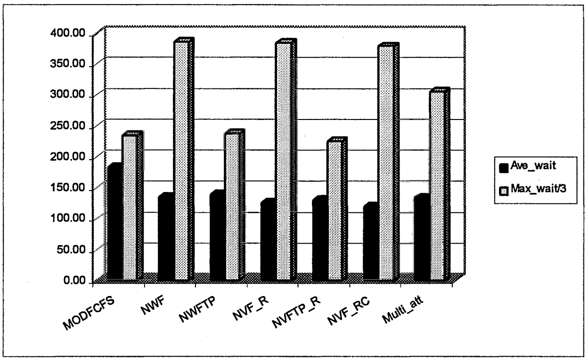


Fig. 8. Performance of dispatching rules when 5 vehicles are used

When five vehicles are used, the vehicle utilization is not too high (about 75%) NVF_R and NVF_RC appear to be best rules according to minimizing the average waiting time criterion. The reason is that, in this case, vehicle unnecessary movements are saved. Table 7 and Fig. 8 indicate that the rules which attempt to re-assign moving vehicles to closer loads have a good performance in average waiting time and the truncation rule NVFTP_R leads to smallest values of the maximum load waiting time.

Table 8. Rank of dispatching rules (Tukey - confidence interval 95%)

Average waiting time						**Max waiting time**			
	Subset						*Subset*		
RULES	1	2	3	4		RULES	1	2	3
NVF_RC	119.81					**NVFTP_R**	**675.39**		
NVF_R	125.99	125.99				MODFCFS	704.01		
NVFTP_R		**130.85**	130.85			NWFTP	710.68		
Multi_att		134.11	134.11			Multi_att		914.82	
NWF		134.19	134.19			NVF_RC			1136.92
NWFTP			139.51			NVF_R			1154.44
MODFCFS				182.46		NWF			1156.99

** In ranking tables, rules are marked as good rules (in Bold) if their average and max waiting time are in first groups for both criteria or one of the performance indicator (average or max waiting time) in the first group and the another in the second group.*

In the performance-ranking table (Table 8), the NVF_RC rule performs significantly well in terms of average waiting time and NWTP_R, NWFTP, MODFCFS perform very well in the max waiting time. When the maximum load waiting time is not very important for the system, the rule with vehicle re-assignment and cancellation is the best. Otherwise the truncation rule with vehicle re-assignment is the better one.

5.1.3 Performance for 6 vehicles case

Table 9. Experimental results for the distribution center (6 vehicles)

6 vehicles							
	MODFCFS	NWF	NWFTP	NVF_R	NVFTP_R	NVF_RC	Multi_att
Avewait	113.64	103.38	103.72	92.50	93.40	88.26	103.58
std	13.22	5.28	5.94	6.07	7.34	5.96	5.59
maxwait	377.87	593.47	427.14	620.18	440.31	575.31	498.09
VehUtil%	66.27	65.20	65.18	62.33	62.64	> 47.10	65.17

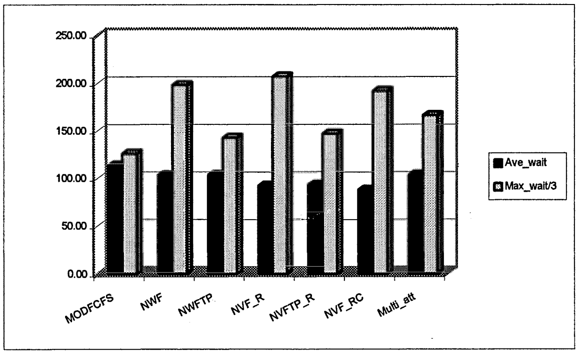


Fig. 9. Performance of dispatching rules when 6 vehicles are used

A very similar result has been obtained for this case compared with the 5 vehicles case. Smaller values of average waiting times are obtained by using rules with vehicle re-assignment. As can be seen in **Table 10**, the NVF_RC rule obtained the smallest value of the average load waiting time. The rules with reassigning of moving vehicles (NVF_R and NWTP_R) perform well also. The MODFCFS rule leads to the smallest value of the maximum load waiting time but is still the worst one in terms of the average waiting time.

Table 10. Rank of dispatching rules (Tukey - confidence interval 95%)

Average waiting time

RULES	Subset 1	Subset 2	Subset 3	Subset 4
NVF_RC	88.26			
NVF_R		92.50		
NVFTP_R		**93.40**		
NWF			103.38	
Multi_att			103.58	
NWFTP			103.72	
MODFCFS				113.64

Max waiting time

RULES	Subset 1	Subset 2	Subset 3
MODFCFS	377.87		
NWFTP	427.14	427.14	
NVFTP_R	**440.31**	440.31	
Multi_att		498.09	
NVF_RC			575.31
NWF			593.47
NVF_R			620.18

The average load waiting time realized by NVFTP_R lies in the second group and the max load waiting time realized by this rule lies in the first group, so NVFTP_R is a very good rule in this case. The smallest value of the average load waiting time in this case is gained by NVF_RC. The max load waiting time obtained by NVF_RC is not far from first group so we can say that this rule is good here as well.

5.2 The Glass Production Plant

Because of the special structure of the glass production plant, in order to keep the average and max load waiting time at acceptable level, we did not experiment with less than seven vehicles. In this case, the workload is quite different between weekdays and weekends. The vehicle utilization is high during weekdays, but is rather low at the weekends. This leads to a low overall vehicle utilization. If we decrease number of used vehicle to less than seven, the average and max load waiting time become very high because of short of a vehicle shortage during weekdays.

5.2.1 Performance for 7 vehicles case

The vehicle utilization is not so high in this case and NVF_RC performs quite well in terms of the average waiting time. The NVFTP_R has a good performance in terms of the max waiting time and this rule performs not so bad in terms of the average waiting time. The Multi_att rule is also a good rule in this case.

Table 11. Experimental results for the glass production plant (7 vehicles)

7 vehicles							
	MODFCFS	**NWF**	**NWFTP**	**NVF_R**	**NVFTP_R**	**NVF_RC**	**Multi_att**
Avewait	611.90	276.61	317.95	270.73	316.68	257.22	277.68
std	92.93	17.67	31.60	24.93	45.36	21.32	21.26
maxwait	3480.55	4394.47	3255.04	4818.29	3042.39	5365.26	4105.21
VehUtil%	59.02	56.12	56.77	55.25	56.17	> 40.94	56.10

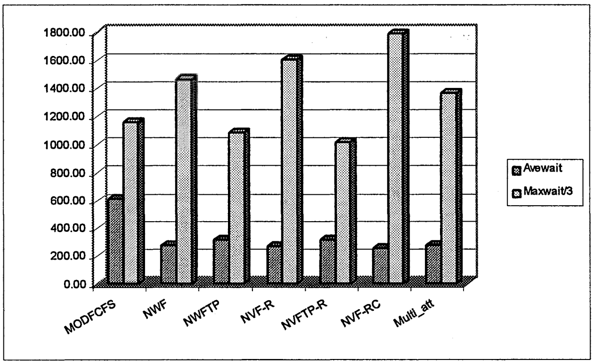


Fig. 10. Performance of dispatching rules when 7 vehicles are used

Table 12. Rank of dispatching rules (Tukey - confidence interval 95%)

Average waiting time					Max waiting time		
	Subset					*Subset*	
RULES	1	2	3		RULES	1	2
NVF_RC	**257.22**				**NVFTP_R**	**3042.39**	
NVF_R	270.73	270.73			**NVFTP**	**3255.04**	3255.04
NWF	276.60	276.60			MODFCFS	3480.55	3480.55
Multi_att	277.68	**277.68**			**Multi_att**	**4105.21**	4105.21
NVFTP_R		**316.68**			NWF	4394.47	4394.47
NVFTP		**317.95**			NVF_R	4818.29	4818.29
MODFCFS			611.90		**NVF_RC**		**5365.26**

5.2.2 Performance for 9 vehicles case

Table 13. Experimental results for the glass production plant (9 vehicles)

9 vehicles							
	MODFCFS	NWF	NWFTP	NVF_R	NVFTP_R	NVF_RC	Multi_att
Avewait	273.32	200.53	204.41	197.21	216.26	185.90	206.87
std	22.31	11.19	6.19	8.92	10.58	6.53	12.51
maxwait	1808.67	2638.81	1501.33	2473.18	1835.38	2066.98	2501.12
VehUtil%	50.25	48.04	48.35	47.44	48.55	> 31.41	48.25

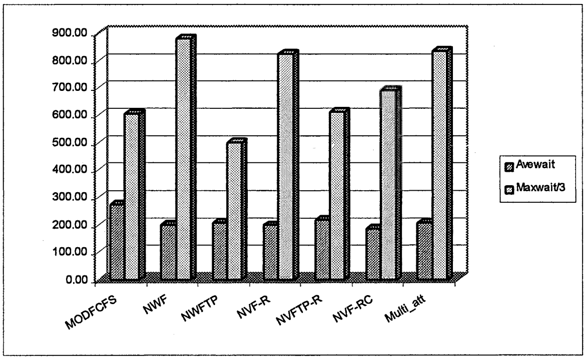


Fig. 11. Performance of dispatching rules when 9 vehicles are used

Table 14. Rank of dispatching rules (Tukey - confidence interval 95%)

Average waiting time				
	Subset			
RULES	1	2	3	4
NVF_RC	**185.90**			
NVF_R	197.21	197.21		
NWF	200.53	200.53		
NWFTP		**204.41**	204.41	
Multi_att		**206.87**	206.87	
NVFTP_R			216.26	
MODFCFS				273.32

Max waiting time		
	Subset	
RULES	1	2
NWFTP	**1501.33**	
MODFCFS	1808.67	1808.67
NVFTP_R	1835.38	1835.38
NVF_RC	**2066.98**	2066.98
NVF_R	2473.18	2473.18
Multi_att	**2501.12**	2501.12
NWF		2638.80

The vehicle utilization is smaller in this case and NVF_RC performs quite well in both performance criteria: the average waiting time and the max waiting time. The NWFTP, NVFTP_R and Multi_att have a good performance in terms of the max waiting time but they perform not so well in terms of the average waiting time. In this case, the NVF_RC is proved to be the best rule.

5.2.3 Performance for 11 vehicles case

Table 15. Experimental results for the glass production plant (11 vehicles)

11 vehicles							
	MODFCFS	NWF	NWFTP	NVF_R	NVFTP_R	NVF_RC	Multi_att
Avewait	209.10	180.80	181.95	175.16	178.20	167.10	180.66
std	9.58	6.06	4.88	5.58	7.14	5.91	6.12
maxwait	1343.28	1577.99	1170.17	1737.66	1266.46	1679.05	1543.97
VehUtil%	44.05	43.03	42.86	41.51	42.55	> 25.15	42.74

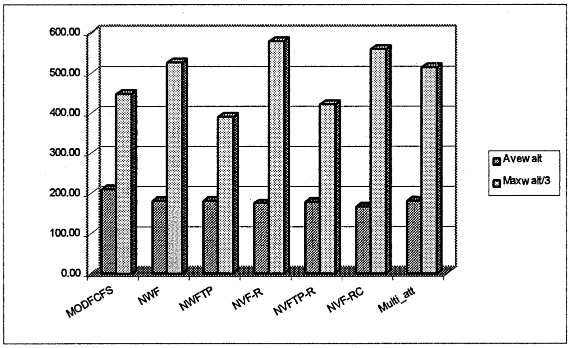


Fig. 12. Performance of dispatching rules when 11 vehicles are used

Table 16. Rank of dispatching rules (Tukey - confidence interval 95%)

Average waiting time					Max waiting time			
	Subset					*Subset*		
RULES	1	2	3		RULES	1	2	3
NVF_RC	**167.10**				**NWFTP**	**1170.17**		
NVF_R		175.16			**NVFTP_R**	**1266.46**	1266.46	
NVFTP_R		**178.20**			MODFCFS	1343.28	1343.28	1343.28
Multi_att		**180.66**			**Multi_att**	**1543.97**	1543.97	1543.97
NWF		**180.80**			**NWF**	**1577.99**	1577.99	1577.99
NWFTP		**181.95**			**NVF_RC**		**1679.05**	1679.05
MODFCFS			209.10		NVF_R			1737.66

Results show that the NVF_RC rule is also a very good rule in this case. This rule performs as the best rule in terms of the average waiting time and performs not so bad in terms of the max waiting time. The NVFTP_R is also a good rule in this case. This rule performs well in both average and maximum load waiting times. The MODFCFS is the worst rule in all cases.

5.3 Discussion

Experimental results indicate that MODFCFS is the worst rule. This rule tends to minimize the max load waiting time, but in our experiment the NWFTP, NVFTP_R and Multi_att also give us a low value for the max load waiting time in comparison with MODFCFS while they lead to a much smaller value of the average load waiting time.

Results from the EDC case show that the system performance is strongly affected by the vehicle utilization (or by number of vehicles) and dispatching rules. If the vehicle utilization is high, rules with vehicle reassignment (and cancellation) seem to be not efficient. Alternatively, the Multi_att and NWFTP rule can keep the average waiting time as small as the corresponding value when NWF is used and can furthermore reduce the maximum load waiting time significantly. When the vehicle utilization is not high, the rules with vehicle reassignment (and cancellation) prove to be good ones. The reason is that, in this case, vehicles may have lots of unnecessary movements (such as moving to a parking place when a load is ready to be picked up), which can be avoided by using rules with vehicle reassignment. The impact of truncation rules (such as NWFTP) on the max load waiting time is less significant when vehicle utilization is low, but still some improvement in the max load waiting time can be gained.

Similar to Bozer and Yen (1996) and Van der Meer (2000), we have found that the NWF_RC rule (MOD STTF in Bozer and Yen, 1996) is efficient in reducing the average load waiting time when the vehicle utilization is not too high. Van der Meer (2000) found that the reassignment rules are not efficient for structured shift arrival jobs (arrival jobs with overlapping between inbound, labeling and outbound jobs - refer to Van der Meer (2000) for more detail). However, in our study the reassignment rules are still efficient when the vehicle utilization is not high. A possible reason is that the experiment layouts for reassignment rules in his research are simple (U and I layouts), so it is not easy to observe the impact of reassignment rules in that case. The Multi_att rule from Klein and Kim (1996) has a good performance in general, particularly when the vehicle utilization is high. However, when the vehicle is not high, dispatching rules using reassignment perform better.

6 Conclusions and Further Research

In this paper, we propose several simple dispatching rules which are applicable in practice. Using simulation, we have evaluated the performance of these rules in two practical cases based on two performance criteria: average and max load waiting times. Results show that these rules are efficient and can help to improve the system performance.

From the analysis, we rank the dispatching based on their performance (Table 17). We have found that, in general, two hybrid rules: NVF_RC and NVFTP_R are the two best overall choices. The Multi_att rule is also a good rule in practice. However, this rule is more complicated and its overall score is a bit lower than NVF_RC and NVFTP_R.

Table 17. Ranking for dispatching rules

Rank	General	High vehicle VehUtil% (> 80%)	Low vehicle VehUtil% (< 80%)
1	NVF_RC, NVFTP_R Multi_att	Multi_att, NWFTP NVFTP_R	NVF_RC, NVFTP_R
2	other rules	other rules	other rules
3	MODFCFS	MODFCFS	MODFCFS

The NVFTP_R and Multi_att rules are recommended for the high vehicle utilization case and the NVFTP_R and NVF_RC are recommended for lower vehicle utilization case. In practice, when vehicles are controlled by human drivers, the rule with vehicle reassignment and cancellation may not be very attractive since a firm schedule is preferred by human drivers and usually it takes sometime for him to react with changes. However, if vehicles are completely automated we can implement this type of rule easily.

For further research, it is necessary to investigate effects of hybrid dispatching rules for other cases to prove the effectiveness of them in general. When information about load arrivals is available some time in advance, a dynamic algorithm to schedule vehicles near optimally can be very useful for improving the system performance. The reason is that our previous dispatching rules are myopic rules and do not take into account future information to dispatch vehicles. Usually more information should lead to a better schedule and a better system performance, although often at the expense of more complex control.

References

Bozer, Y. A. / Yen, C. (1996): Intelligent dispatching rules for trip-based material handling systems. *Journal of Manufacturing Systems,* 15(4), 226-239.

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.

Egbelu, P. J. / Tanchoco, J. M. A. (1984): Characterization of automated guided vehicle dispatching rules. *International Journal of Production Research,* 22(3), 359-374.

Fleischmann, B. / Gnutzmann, S. / Sandvoß, E. (2003): Dynamic vehicle routing based on on-line traffic information. *Transportation Science (To appear).*

Klein, C. M. / Kim, J. (1996): AGV dispatching. *International Journal of Production Research,* 34(1), 95-110.

Srinivasan, M. M. / Bozer, Y. A. / Cho, M. (1994): Trip-based material handling systems: Throughput capacity analysis. *IIE Transactions,* 26(1), 70-89.

Van der Meer, J. R. (2000): Operational control of internal transport system. Erasmus University Rotterdam.

Van der Meer, J. R. / De Koster, R. (2000): Centralized internal transport systems analyzed and classified. In: Progress in Material Handling Research, Ann Arbor Michigan.

Chapter 4

Reverse Logistics

Travel Distance Estimation in Single-block ABC-Storage Strategy Warehouses

Tho Le Duc and René de Koster

Erasmus University Rotterdam, Rotterdam School of Management/ Faculteit Bedrijfskunde, Rotterdam, The Netherlands

Abstract. Order picking has been considered as the most critical operation in warehousing. Four mayor components that greatly affect the efficiency of the operation are the routing method, the storage strategy, the warehouse layout and the batching policy. While the routing in warehouse problem, to some extends, has been investigated thoroughly, the others have not been widely studied. One of the essential problems that need to be solved when dealing with these problems is the problem of estimating the average travel distance of a picking tour.

This paper deals with the problem of estimating the average travel distance in picker-to-parts narrow-storage-aisle ABC-storage strategy warehouses. We first propose a formula that enables us to estimate the average travel distance within a storage-aisle. Applying this formula, we then derive formulas for the average travel distance of a pick tour in a typical multiple-aisle warehouse. The routing policy used is the return heuristic, meaning that order pickers only enter storage-aisles containing picks and they enter and leave the storage-aisle from the same end. To test the quality of the formulas we built several simulation models. We found that the differences between results obtained from simulation and analytical models were negligible. The findings provide a concrete foundation for two promising optimization problems: storage zone boundary and layout optimization problem.

Keyword: Order picking, storage strategy, travel distance estimation

1 Introduction

Order picking can be described as the process of retrieving a number of items from their storage locations to fill a number of independent customer orders. Although it seems to be a simple operation, it is a laborious process and a major cost factor for warehouses in both production and distribution environments. It has been estimated that the cost of order picking may contribute up to 55% of all operating costs in a typical warehouse (see, for example, Tompkins et al., 1996). Furthermore, new developments and trends in the global economy have had a substantial impact on the order picking process; the demand pattern has changed from few but large orders to many but small orders, which have to be processed within very tight time windows. Therefore, simultaneously reducing the cost and increasing the speed of order picking becomes more and more critical.

B. Fleischmann et al. (eds.), *Distribution Logistics*

The performance and efficiency of the picking operation mainly depends on the following factors (see also Peterson and Schmenner, 1999).

1. The demand pattern.
2. The configuration (layout) of the warehouse.
3. The storage strategy: how to allocate items within the warehouse (e.g. randomly assigning items to storage locations or assigning items based on their order frequencies, picking volume or cube-per-order index (COI), etc.).
4. The batching method: how to group orders and divide orders among order pickers.
5. The routing and sorting method: how to determine the sequence of the items to be picked and how to consolidate those items.

Except for the first factor, the others are usually controllable and have been investigated by many researchers recently. The routing method is the first issue that has received considerably attention in the warehousing literature. Some prominent papers on the subject are Ratliff and Rosenthal (1983), Goetschalckx and Ratliff (1988), Hall (1993), Petersen (1997), De Koster and Van der Poort (1998), De Koster et al. (1999[a]), Roodbergen and De Koster (2001) and Roodbergen (2001). Theory covering the layout design and the interaction between routing and layout can be found in, for example, Jarvis and McDowell (1991), Petersen and Schemenner (1999) and Roodbergen (2001). Batching issues are mentioned in Gibson and Sharp (1992), Pan and Liu (1995), Rosenwein (1996), Tang and Chew (1997), De Koster et al. (1999[b]) and Gademann et al. (2001). It is quite obvious that with a given demand pattern, a routing and a batching policy, a dedicated storage strategy results in a shorter total travel distance than to the random strategy. Even so, the research on storage assignment issues is fairly limited. Gibson and Sharp (1992), Gay et al. (1992), Petersen and Schmenner (1999) and Dekker et al. (2002) discuss a volume-based strategy that is locating high volume items close to the depot. Caron et al. (1998) mention a COI-based storage strategy, where items are located based on their ratio of the required storage space to the order frequency.

In this paper we focus on picker-to-parts, narrow-aisle, manual-operation order picking systems. The employed storage strategy is ABC, meaning that the items are divided in classes (for example A, B, C, ...) based on their pick frequencies. The locations are divided in an equal number of classes such that the most fast moving items are assigned to the locations closest to the depot. The ABC-storage strategy is widely used in practice because it is convenient to implement and maintain; it can easily handle assortment changes or changes in pick frequency. In addition, using the ABC-storage strategy leads to a substantial reduction in order pick travel time as compared to the random storage (see Hausman 1976). The average travel distance of a picking tour, which is linearly related to the travel time, is crucial: it is often used as the objective function for warehouse optimization problems. Therefore, our aim was to provide an analytical model for estimating the average travel distance of a picking tour in this type of warehouses.

Several papers exist that address the problem of distance estimation in warehouses. Some of them are, for example, Ashayeri et al. (2001), Bozer and

White (1984), Sarker and Babu (1995) and Chew and Tang (1999). Most of these papers focus on single-aisle unit-load, automatic storage and retrieval systems, only few of them involve conventional multiple-aisle order-picking systems. Recently, Roodbergen (2001) proposed formulas for estimating the average travel distance in both single-block and multi-block conventional warehouses under the random storage strategy (incoming products are located randomly over the warehouse) and traversal routing strategy. His formulations perform better than previously existing formulas. The limitation of these formulations is that they cannot be applied to non-random (i.e. ABC, COI-based) storage strategy warehouses. Caron et al. (1998,2000) discuss the COI-based storage strategy and propose travel time models for estimating the average travel distance of a tour when either traversal or return routing heuristics are employed. The COI-based strategy assigns products with a low ratio of the required storage space to the order frequency to the locations nearest to the depot. This storage strategy differs from the ABC-storage strategy, because they use different assignment criteria. The ABC-storage strategy assigns products to storage locations in groups while the COI-based storage strategy, mentioned in Caron et al. (1998,2000), allocates products on an individual basis. In practice, such a COI-based strategy can only be applied to very stable assortments with limited changes in order frequency and limited changes in the stored volume. This situation is, however, more and more uncommon in practice.

The following notations are generally used:

- a total number of storage-aisles (or pick-aisles).
- l_{ij} partial length of storage-aisle j used for storing of class i.
- q number of picks (or order-lines) in a picking tour (the pick-list size).
- c number of (product) classes.
- L length of a storage-aisle.
- w_a width of the cross-aisle.
- w_b centre-to-centre distance between two consecutive storage-aisles.
- P_i order frequency of class i.
- p_{ij} probability that an item of class i located in storage-aisle j is ordered.
- TD^{CA} total travel distance within the 'cross-aisle'.
- TD^{WA} total travel distance within the 'within-aisle'.

2 Average Travel Distance within a Storage-aisle

In this section, we consider a single aisle (aisle j) with a configuration given in figure 2. Zone 1, zone 2, ... and zone c are reserved for items of class 1, 2, ...and c respectively. We assume that within each zone items are uniformly distributed.

By conditioning on the farthest location of the requested items, the expected time from the starting point (see figure 1) to the farthest pick location to pick up q_j picks can be computed as follows:

$$D_j(q_j,c) = \sum_{i=1}^{c} P_{ij}\mathrm{E}(\textit{travel time} \mid \textit{farthest pick in zone } i) = \sum_{i=1}^{c} P_{ij}d_i \ . \qquad (1)$$

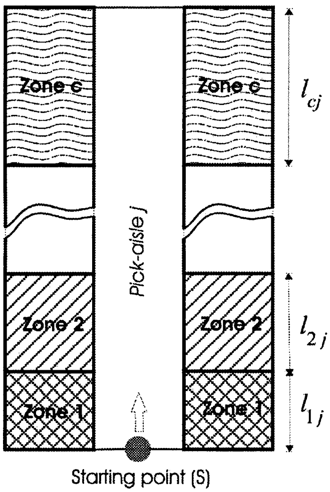


Fig. 1. Single-aisle layout

From (1), we can see that in order to determine $D_j(q_j,c)$ (the expected travel distance from the starting point to the farthest pick location) we have to determine P_{ij} (the probability that the farthest pick is in zone i) and d_{ij} (the corresponding expected one-way travel distance). We consider the following situations:

- the farthest pick is in zone 1,
- the farthest pick is in zone 2 and
- the farthest pick is in zone i $(i=3..c)$.

2.1 Farthest pick is in zone 1

If the farthest pick is in the zone 1, this means that all picks are in zone 1 and no pick is in the zones from 2 to c:

$$P_{1j} = p_{1j}^{\ q_j} \qquad (2)$$

$$d_{1j} = \mathrm{E}(\textit{travel distance} \mid \textit{all picks in zone 1}) = \frac{l_{1j}q_j}{q_j+1} \qquad (3)$$

(3) is based on the well-known property that the expectation of the maximum of q_j continuous uniformly distributed [0,1] variables equals $\frac{q_j}{q_j+1}$. Recall that p_{ij} $\left(\sum_{i=1}^{c} p_{ij} = 1\right)$ is the probability that an item of class i located in the aisle j is ordered.

2.2 Farthest pick is in zone 2

The farthest pick in zone 2 is equivalent with at least one pick in zone 2 and no pick in the zones from 3 to c (or all picks in zone 1 and zone 2 but not all picks in zone 1):

$$P_{2j} = \left(p_{1j} + p_{2j}\right)^{q_j} - \left(p_{1j}\right)^{q_j}. \tag{4}$$

$$d_{2j} = l_{1j} + l_{2j}\mathrm{E}\left(\frac{N_2}{N_2+1}\right). \tag{5}$$

Where N_2 is the number of picks in zone 2. It is rather difficult to compute d_{2j} based on (5). Therefore, we estimate d_{2j} as follows. First, we calculate $\mathrm{E}\left(N_2\right)$, the expected number of picks in zone 2:

$$\mathrm{E}\left(N_2\right) = \sum_{n=1}^{q_j} nP(n\ \textit{picks in zone 2} \mid \textit{all picks in zones 1\&2, and not all in zone 1})$$

$$= \sum_{n=1}^{q_j} n \frac{P(n\ \textit{picks in zone 2, all picks in zones 1\&2, and not all in zone 1})}{P(\textit{all picks in zones 1 and 2, and not all picks in zone 1})}$$

$$= \sum_{n=1}^{q_j} n \frac{P(n\ \textit{picks in zone 2 and } (q_j\text{-}n) \textit{ picks in zone 1})}{P_{2j}}$$

$$= \frac{\sum_{n=1}^{q_j} n \binom{q_j}{n} \left(\frac{p_{2j}}{p_{1j}+p_{2j}}\right)^{n} \left(\frac{p_{1j}}{p_{1j}+p_{2j}}\right)^{q_j-n}}{(p_{1j}+p_{2j})^{q_j} - p_{1j}^{\ q_j}} = \frac{\frac{q_j p_{2j}}{p_{1j}+p_{2j}}}{(p_{1j}+p_{2j})^{q_j} - p_{1j}^{\ q_j}}.$$

The last step is based on the property of Binomial distribution. Then, d_{2j} can be estimated as follows:

$$d_{2j} \approx l_{1j} + l_{2j}\frac{\mathrm{E}\left(N_2\right)}{\mathrm{E}\left(N_2\right)+1}$$

$$= l_{1j} + \frac{\dfrac{l_{2j}\dfrac{q_j p_{2j}}{p_{1j}+p_{2j}}}{(p_{1j}+p_{2j})^{q_j} - p_{1j}^{\ q_j}}}{\dfrac{\dfrac{q_j p_{2j}}{p_{1j}+p_{2j}}}{(p_{1j}+p_{2j})^{q_j} - p_{1j}^{\ q_j}} + 1} = l_{1j} + \frac{\dfrac{l_{2j} q_j p_{2j}}{p_{1j}+p_{2j}}}{\dfrac{q_j p_{2j}}{p_{1j}+p_{2j}} + (p_{1j}+p_{2j})^{q_j} - p_{1j}^{\ q_j}}$$

$$= l_{1j} + \frac{l_{2j} q_j p_{2j}}{q_j p_{2j} + \left(p_{1j}+p_{2j}\right)\left[(p_{1j}+p_{2j})^{q_j} - p_{1j}^{\ q_j}\right]}. \quad (6)$$

2.3 Farthest pick is in zone i $(i=3..c)$

If farthest pick is in zone i $(i=3..c)$ then we can apply the same procedure as in two previous situations, we have:

$$\underset{i\geq 3}{P_{ij}} = \left(p_{1j} + p_{2j} + \ldots + p_{ij}\right)^{q_j} - \left(p_{1j} + p_{2j} + \ldots + p_{i-1,j}\right)^{q_j} \quad (7)$$

$$\underset{i\geq 3}{d_{ij}} \approx \sum_{k=1}^{i-1} l_{ij} + \frac{l_{ij} q_j p_{ij}}{q_j p_{ij} + P_{ij} \sum_{k=1}^{i} p_{kj}}. \quad (8)$$

Finally, substituting (2), (3), (4), (6), (7) and (8) into (1), we obtain:

$$D_j(q_j, c) \approx p_{1j}^{\ q_j} \frac{l_{1j} q_j}{q_j + 1} + \sum_{i=2}^{c} \left[P_{ij}\left(q_j\right) \left(\sum_{k=1}^{i-1} l_{kj} + \frac{l_{ij} q_j p_{ij}}{q_j p_{ij} + P_{ij}\left(q_j\right) \sum_{k=1}^{i} p_{kj}} \right) \right], \quad (9)$$

where $P_{ij}\left(q_j\right) = \left(\sum_{k=1}^{i} p_{kj}\right)^{q_j} - \left(\sum_{k=1}^{i-1} p_{kj}\right)^{q_j}$.

2.4 Simulation test

To test the quality of formula (9), we built a simulation model in Microsoft Excel using VBA (Visual Basic for Applications). In the test, we considered the case that there were only three classes. The 'three classes' (namely A, B and C) storage strategy is the most popular one in practice.

In the case of three classes, it is easy to verify that:

$$D(g,3) \approx p_A{}^g \frac{l_A g}{g+1} + \left\{(p_A + p_B)^g - (p_A)^g\right\} \cdot$$

$$\left\{l_a + \frac{l_B g p_B}{g p_B + (p_A + p_B)\left[(p_A + p_B)^g - p_A{}^g\right]}\right\}$$

$$+ \left\{1 - (p_A + p_B)^g\right\}\left\{l_A + l_B + \frac{l_C g p_C}{g p_C + 1 - (p_A + p_B)^g}\right\}, \qquad (10)$$

where g is the number of picks. p_A, p_B, p_C and l_A, l_B, l_C are the order frequencies and storage length of class A, B and C correspondingly.

We considered three different ABC-storage assignments (ABC curves), namely: skewed, random and medium. The skewed assignment means frequently ordered items occupy only a very small portion of the total storage space. The random assignment means that no distinction between items classes in the term of order frequency and required space; they are randomly located within the warehouse. Finally, the medium assignment is in between two above-mentioned patterns. The percentages of assigned space and order frequencies of classes are listed on table 1. The effective pick-list size varies from 4 to 60 items per picking tour.

Table 1. Storage assignment schemes

Assignment	A-class	B-class	C-class
Skewed	20/70	30/20	50/10
Medium	30/50	30/30	20/40
Random	33.33/33.33	33.33/33.33	33.33/33.33

For each scenario, a simulation model was run 10000 times, after which the 98% confidence interval for the average travel distance was determined. Figure 2 shows the details of the results, where the length of the storage-aisle was normalized to 1. The differences are generally less than 5%. It confirms that the formula provides a good approximation for the travel distance in a single aisle. When the pick-list size is very large, the difference is very small. It is because the order picker has to travel almost the entire storage-aisle to pick up a large number of items; the average travel distances found by both methods are close to the maximum travel distance (distance to travel the entire warehouse). Furthermore, the results from the approximation are always higher than the corresponding simulation results. The reason is that we overestimate the conditional expected travel distance d_{ij}.

With respect to the storage assignment, firstly, the ABC-storage can lead to a saving up to 32% and 43% compared to the random strategy for the medium and skewed assignment. This finding is in line with the results in previous research on the ABC-storage assignment (see, for example, Hausman et al. (1976); Graves et al. (1977); Schwarz et al. (1978)) . Secondly, the formula is more favorable for the random compared to the medium and skewed assignment. In the worst case, the difference between the approximation and simulation for the random, skewed and medium assignment are 2.47%, 4.55% and 4.57% respectively. The reason is that the variation is higher in the cases of the skewed and medium assignment.

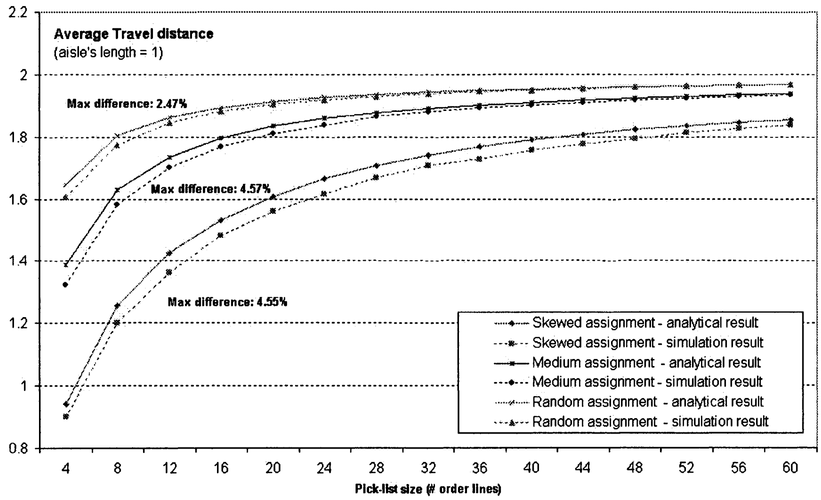


Fig. 2. The difference between approximation and simulation results, $L = 1$.

3 Average Travel Distance in Multiple-aisle Layouts

In this section, we will apply the formula that we have derived in the previous section to a multiple-aisle warehouse with one cross-aisle (right in the middle) and multiple storage-aisles. Figure 2 shows the layout when the number of storage-aisles is 4. This layout stems from Caron et al. (1998) and is fairly common in practice. The order picking process in multiple-aisle layouts can be described as follows. Starting from the depot, the picker enters the nearest storage-aisle that contains picks (called requested storage-aisle). Then the picker travels in the storage-aisle (goes as far as the farthest pick location), picks up the requested items, and returns to the middle of the cross-aisle. From that point, the picker moves to the next requested storage-aisle, which is nearest to the current position. After visiting the farthest requested storage-aisle, he returns to the starting point. One may realize that the routing method used here is the return heuristic. In literature, there exist many other routing methods (see for example, Ratliff and Rosenthal 1983 and De Koster and Van der Poort 1998). However, only the return heuristic is relevant for this type of layout. It is noted that it does not matter whether the order picker completely serves one side of the warehouse first and then goes to the other side or simultaneously serves both sides, because in both cases we have exactly the same total travel distance.

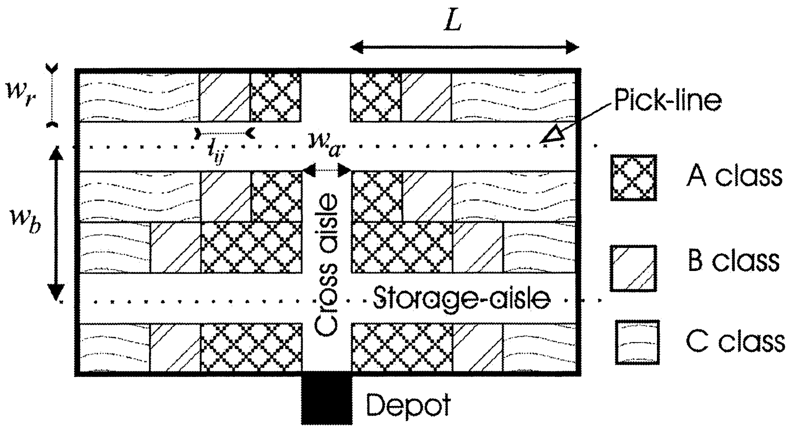


Fig. 3. Multiple-aisle layout

Suppose that products are classified into c classes and the pick-list consists of q items. Let $l_{ij} \geq 0$ denote the length in storage-aisle $j\ (j = 1..a)$ used for storing class $i\ (i = 1..c)$ and $P_i > 0$ denotes the order frequency of class $i\ (i = 1..c)$. We assume that l_{ij} and P_i are given. Furthermore, the number of picks requested from a class is proportional to the order frequency of the class.

We can see that the travel distance of a picking tour consists of two components: 'within-aisle' and 'cross-aisle' travel distance. In the following, we will step by step estimate these quantities.

3.1 Estimate for the within-aisle travel distance

Let p_{ij} be the probability that an item belonging to class i located in storage-aisle j is ordered. We have:

$$p_{ij} = P_i \frac{l_{ij}}{\sum_{j=1}^{a} l_{ij}} \quad \forall (i = 1..c, j = 1..a).$$

Clearly, if one item is ordered then the probability that storage-aisle j is visited is $\sum_{i=1}^{c} p_{ij}$. Thus, if q items are required then the probability that the storage-aisle j is visited is:

$$m_j = 1 - \left(1 - \sum_{i=1}^{c} p_{ij}\right)^q .$$

The average travel distance from the center line of the cross-aisle to the farthest pick in storage-aisle j (given that the storage-aisle j is visited) is:

$$k_j = w_a/2 + D_j(q_j, c).$$

Where q_j is the conditionally expected number of items to be picked from the storage-aisle j:

$$q_j = \frac{\text{expected number of picks in aisle } j}{\text{probability that aisle } j \text{ is visited}} = \frac{q\sum_{i=1}^{c} p_{ij}}{m_j}.$$

And, $D_j(q_j, c)$ is the average travel distance from the starting point to the farthest pick in the storage-aisle j to pick up q_j items. Applying formula (9), we have:

$$D_j(q_j,c) \approx p_{1j}'^{\,q_j} \frac{l_{1j}q_j}{q_j+1} + \sum_{i=2}^{c}\left(P_{ij}\left(q_j\right)\left[\sum_{k=1}^{i-1} l_{kj} + \frac{l_{ij}q_j p'_{ij}}{q_j p'_{ij} + P_{ij}\left(q_j\right)\sum_{k=1}^{i} p'_{kj}}\right]\right),$$

where $p'_{ij} = \dfrac{p_{ij}}{\sum_{i=1}^{c} p_{ij}}$ $\left(\forall j = 1..a\right)$. It is easy to verify that $\sum_{i=1}^{c} p'_{ij} = 1$ $\left(\forall j = 1..a\right)$.

The average 'within-aisle' travel distance of a tour can be now computed as follows

$$TD^{WA} \approx 2\sum_{j=1}^{a}\left[\left(\text{expected travel distance if } j \text{ is visited}\right) * \left(\text{prob. that } j \text{ is visted}\right)\right]$$

$$= 2\sum_{j=1}^{a} k_j m_j = 2\sum_{j=1}^{a}\left\{\left[w_a/2 + D_j(q_j,c)\right]\left[1 - \left(1 - \sum_{i=1}^{c} p_{ij}\right)^q\right]\right\}$$

$$= 2\sum_{j=1}^{a}\left\{ w_a/2 + \left(p'_{1j}\right)^{q_j} \frac{l_{1j}q_j}{q_j+1} + \sum_{i=2}^{c}\left\{ P_{ij}\left(q_j\right)\left(\sum_{k=1}^{i-1} l_{kj} + \frac{l_{ij}q_j p'_{ij}}{q_j p'_{ij} + P_{ij}\left(q_j\right)\sum_{k=1}^{i} p'_{kj}} \right) \right.\right.$$

$$\left.\left. \cdot\left[1-\left(1-\sum_{i=1}^{c} p_{ij}\right)^{q}\right]\right\}\right\} \qquad (11)$$

3.2 Estimate for the cross-aisle travel distance

Given that q picks are required, let n_j $(j = 1..a/2)$ be the probability that at least one of storage-aisles j and $(a-j+1)$ is visited. Due to the symmetrical property of the layout, the probabilities that the storage-aisle j and storage-aisle $(a-j+1)$ is visited are equal. Therefore, n_j $(j = 1..a/2)$ can be determined as follows:

$$n_j = 1-\left(1-m_j\right)\left(1-m_{a-j+1}\right) = 1-\left(1-m_j\right)^2 = 1-\left(1-\sum_{i=1}^{c} p_{ij}\right)^{2q}$$

If we consider the cross-aisle as a 'storage-aisle' and the storage-aisles and the corresponding storage racks around them as storage zones of the 'storage-aisle' then the average cross-aisle travel distance can also be calculated by using formula (9):

$$TD^{CA} \approx 2\left(n'_1\right)^q \frac{\left(w_b/2\right)q}{q+1} + 2\sum_{i=2}^{a/2}\left\{ P_{ij}(q)\left[\sum_{j=1}^{i-1} w_b + \frac{\left(w_b/2\right)qn'_i}{qn'_i + P_{ij}(q)\sum_{j=1}^{i} n'_j} \right]\right\}, \qquad (12)$$

where $n'_j = n_j \Big/ \sum_{j=1}^{a/2} n_j$ $(\forall j = 1..a/2)$. It is noted that $w_b/2$ is the distance that an order picker has to travel along the cross-aisle in order to reach the central line of the follow-up storage-aisle.

Summing up TD^{CA} and TD^{WA} together we obtain a formula for estimating the average travel distance of a picking tour TD.

3.3 Simulation test

We used also simulation to examine the performance of the proposed formula. In the experiment, we considered three warehouses: with 6, 10 and 16 storage-aisles. We also assumed that items were grouped into 3 classes and assigned to storage locations by either the skewed, medium or random assignment mentioned in section 2.1. The effective pick-list size varies from 4 to 60 items per picking tour. In total, we have 135 scenarios to simulate, details about input parameters can be found in figure 3-5 (note that the storage-aisle length is normalized to 1).

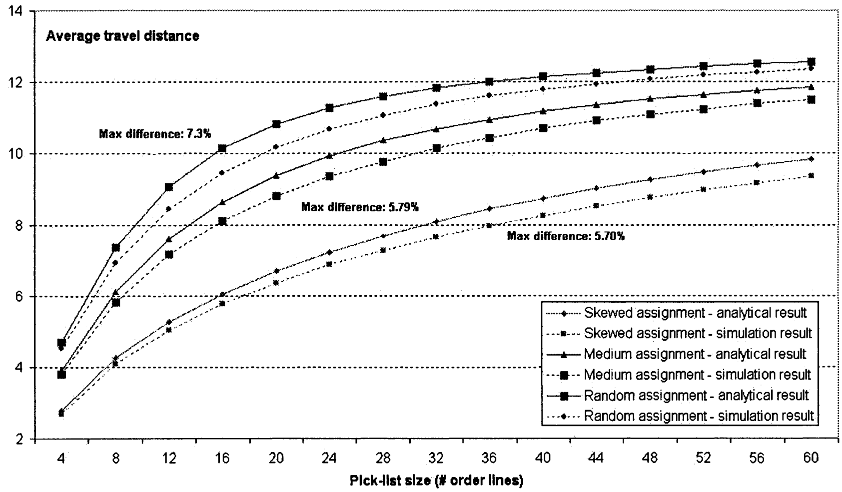


Fig. 4. Differences between approximation and simulation results for the 6-aisle layout with $L = 1, w_a = 0.107, w_b = 0.179$.

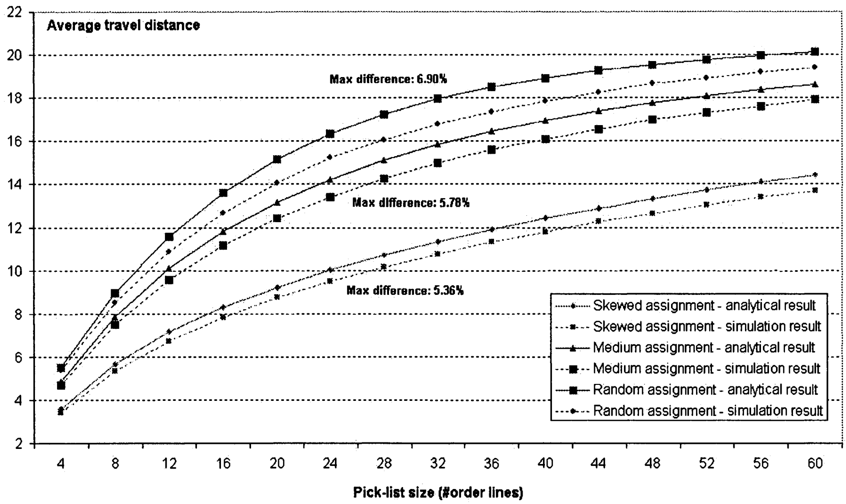


Fig. 5. Differences between approximation and simulation results for the 10-aisle layout with $L = 1, w_a = 0.107, w_b = 0.179$.

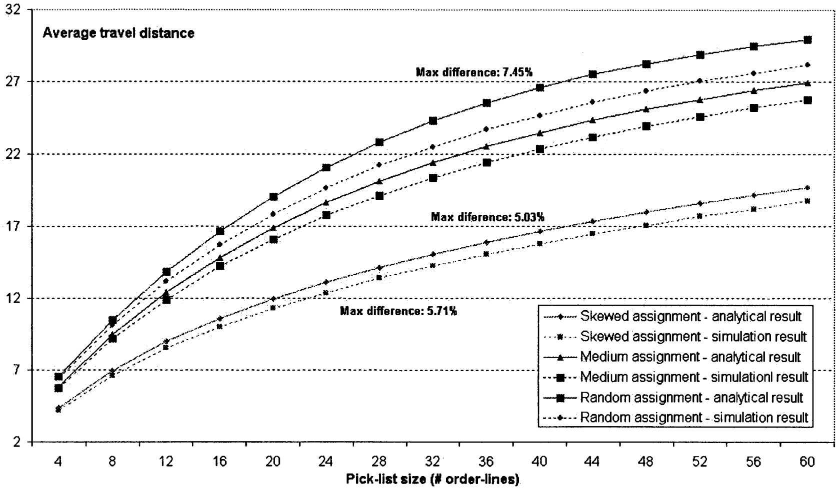


Fig. 6. Differences between approximation and simulation results for the 16-aisle layout with $L = 1, w_a = 0.107, w_b = 0.179$.

For each simulation run we drew q picks (listed in the first column), which were then assigned to zones based on the order frequencies of item classes and the storage spaces for a certain item class in each storage-aisle. Next, within each zone, items were randomly stored. In the comparison, we used 10000 replications for each value of q. Figures 3-5 show the approximation and simulation results for all simulated scenarios. The average travel distance is an increasing concave function of the pick-list size. It reaches the maximum when the pick-list size is very large. Figure 6 delineates the shape of the difference in the case of the layout with 16 storage-aisles. From the figures, we can see that the difference between approximation and simulation results increases when the pick density – defined as the average number of picks (or order-lines) per storage-aisle – increases. It reaches a maximum and then from there it decreases. When the pick-list size is very large, the difference is very small. We can explain this behavior as follows. The difference (or error - ζ) consists of the following two components:

The accumulated error resulting from estimating the average travel distance within all visited storage-aisles (ζ_a). This amount is proportional to the expected number of visited storage-aisles ($\bar{n}$) and the error in estimating the travel distance in a single aisle (ζ_b).

The error in estimating the cross-aisle travel distance (ζ_c). This depends on $\bar{n}$ only.

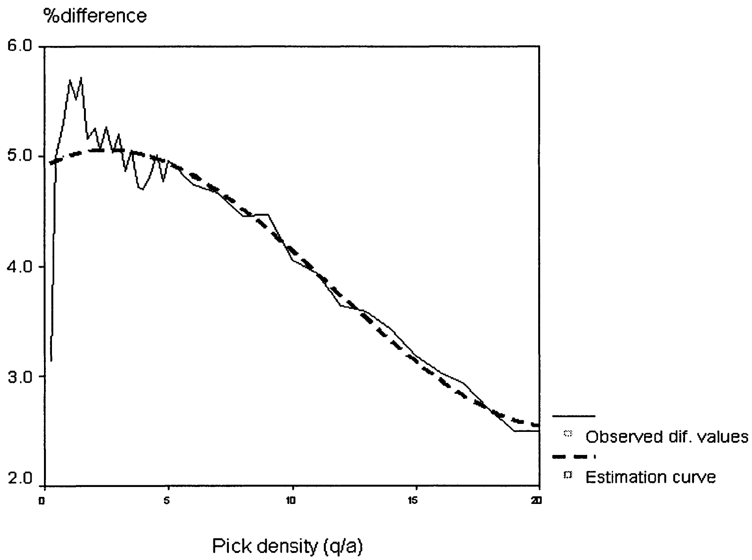


Fig. 7. The effect of the pick density (16-aisle layout)

When the pick-list size increases from a small value, both $\bar{n}$ and ζ_b increase. As a result, ζ_a and ζ_c, and therefore, ζ increase. When the pick-list size is substantially large, $\bar{n}$ reaches the maximal possible number of visited storage-aisles (n) and ζ_c stops increasing. For even larger pick-list sizes ζ_b will start to decrease, since the order picker has to visit the entire warehouse. As a result ζ will decrease as well. The estimate also approaches the maximum travel distance in this case. In accordance with our expectation, the random assignment always provides the longest average travel distance.

4 Concluding Remarks

We have presented an analytical approach for estimating the average travel distance of a picking tour in a typical ABC-storage strategy warehouse. Small differences as compared to simulation outcomes ensure the high accuracy level of the formulas. The analytical approach is very robust; based on it, we may expect to estimate the average travel distance in more complicated ABC-storage strategy layouts.

This study creates a foundation for the following optimization problems.

- *Storage zone optimization problem:* that is the problem of determining how much space should be used for a certain item class in each aisle such that the average travel distance of a picking tour is minimized. In other words, given the order pattern (the order frequencies of item classes), the pick-list size and the layout configuration (the number of storage-aisles, the dimension of aisles), we have to determine the optimal zone boundaries in each aisle with respect to minimizing the average travel distance.

- *Warehouse layout optimization problem*: that is the problem of determining the optimal layout (dimension of the layout, the number of storage-aisles, and the optimal zone boundaries for each aisle) such that the average travel distance of a picking tour is minimized. It differs from the first problem because of the fact that now the number of storage-aisles becomes a decision variable.

In the paper, we have considered a simple layout and for this layout only the return heuristic is relevant. In practice, more complex layouts (thus other routing methods) may be used. The most efficient routing heuristic known in literature is the combined heuristic, where the return and traversal strategy are combined (see Roodbergen, 2001). We think that is an interesting direction for our further research.

References

Ashayeri, J. / Heuts, R. / Valkenburg, M. W. T. / Veraart, H. C. / Wilhelhm, M. R. (2001): A geometric approach to computing expected cycle times for ABC-storage layout in AS/RS. Discussion paper, CentER, Tilburg University, The Netherlands.

Bozer, Y. A. / White, J. A. (1984): Travel-time models for automated storage/retrieval systems. *IIE Transactions* 16, 329-338.

Caron, F. / Marchet, G. / Perego, A. (1998): Routing policies and COI-based storage policies in picker-to-part systems. *International Journal of Production Research* 36(3), 713-732.

Caron, F. / Marchet, G. / Perego A. (2000): Optimal layout in low-level picker-to-part systems. *International Journal of Production Research* 38(1), 101-117.

Chew, E. P. / Tang, L. C. (1999): Travel time analysis for general item location assignment in a rectangular warehouse, *European Journal of Operational Research* 112, 582-597.

De Koster, R. / Van der Poort, E. S. (1998): Routing order pickers in a warehouse: a comparison between optimal and heuristic solutions. *IIE Transactions* 30, 469-480.

De Koster, R. / Van der Poort, E. S. / Wolters, M. (1999a): Efficient orderbatching methods in warehouses. *International Journal of Production Research* 37(7), 1479-1504.

De Koster, R. / Van der Poort, E. S. / Roodbergen, K. J. (1999b): When to apply optimal or heuristic routing of order pickers, In B. Fleischmann et al. (eds.): Advances in Distribution Logistics, Springer Verlag, Berlin, 375-401.

Dekker, R. / De Koster, R. / Van Kalleveen, H. / Roodbergen, K. J. (2002): Quick response practices at the warehouse of ANKOR. *ERIM Report Series Research in Management*, Erasmus University Rotterdam. (Available online: http://www.eur.nl/WebDOC/doc/erim/erimrs20020222120221.pdf).

Gademann, A. J. R. N. / Van der Berg, J. P. / Van der Hoff, H. H. (2001): An order batching algorithm for wave picking in a parallel-aisle warehouse. *IIE Transactions* 33, 385-398.

Gibson, D. R. / Sharp, G. P. (1992): Order batching procedures. *European Journal of Operations Research* 58(1), 57-67.

Goetschalckx, M. / Ratliff, H. D. (1988): Order picking in an aisle. *IIE Transactions* 20(1), 53-62.

Hall, R. W. (1993): Distance approximations for routing manual pickers in a warehouse. *IIE Transactions* 25(4), 76-87.

Hausman, W. H. / Schwarz, L. B. / Graves, S. C. (1976): Optimal storage assignment in automatic warehousing systems. *Management Science* 22(6), 629-638.

Jarvis, J. M. / McDowell, E. D. (1991): Optimal product layout in an order picking warehouse. *IIE Transactions* 23(1), 93-102.

Pan, C. H. / Liu, S. Y. (1995): A comparative study of order batching algorithms. *Omega International Journal of Management Science* 23(6), 691-700.

Petersen, C. G. (1997): An evaluation of order picking routing policies. *International Journal of Operations & Production Management* 17(11), 1098-1111.

Petersen, C. G. / Schmenner, R. W. (1999): An evaluation of routing and volume-based storage policies in an order picking operation. *Decision Sciences* 30(2), 481-501.

Ratliff, H. D. / Rosenthal, A. S. (1983): Orderpicking in a rectangular warehouse: a solvable case of the travelling salesman problem. *Operations Research* 31(3), 507-521.

Roodbergen, K. J. (2001): Layout and routing methods for warehouses. Ph.D. thesis, Erasmus Research Institute of Management, Erasmus University of Rotterdam, The Netherlands.

Roodbergen, K. J. / De Koster, R. (2001): Routing order- pickers in a warehouse with a middle aisle. *European Journal of Operational Research 133(1), 32-43.*

Rosenwein, M. B. (1996): A comparison of heuristics for the problem of batching orders for warehouse selection. *International Journal of Production Research* 34(3), 657-664.

Sarker, B. R. / Babu, P. S. (1995): Travel time models in automated storage/retrieval systems: A critical review. *International Journal of Production Economics* 40, 173-184.

Tompkins, J. A. / White, J. A. / Bozer, Y. A. / Frazelle, E. H. Tanchoco, J. M. A. / Trevino, J. (1996): Facilities Planning (2nd edition). John Wiley & Sons, New York.

The Impact of Constraints in Closed-loop Supply Chains: The Case of Reusing Components in Product Manufacturing

Roland Geyer [1], Luk N. Van Wassenhove [2]

[1] Bren School of Environmental Science and Management, University of California, Santa Barbara, USA
[2] INSEAD, Fontainebleau, France

Abstract. Closed-loop supply chains for end-of-life products are circular production structures that recover products after the end of their useful lives and reprocess them into secondary resources for the original supply chain. Since most products are not designed for end-of-life value recovery closed-loop supply chains are likely to suffer from technical or economic bottlenecks, called constraints. From a modelling perspective closed-loop supply chains are feedback loops, and their constraints can therefore introduce non-linearities into the production system. This chapter discusses two such constraints, limited component durability and finite product life cycles, in detail and quantifies their impacts in the case of value recovery based on component reuse. The analysis shows that these constraints can have a significant impact on the performance of closed-loop supply chains.

1 Introduction

The systems approach in supply chain management is based on the fundamental result that optimising system components generally leads to sub-optimal solutions for the system as a whole (von Bertalanffy 1976). To avoid this, supply chain management seeks to optimise supply chains across its economic activities and agents. In the traditional view these supply chain activities do not include anything beyond the sale and delivery of the final good. (Lee & Billington 1995) define a supply chain as *[...] a network of facilities that procure raw materials, transform them into intermediary goods and then final products, and deliver the products to customers through a distribution system.* Therefore most supply chain models are developed to manage and optimise linear flows of raw materials and goods from suppliers via producers to final customers (Tayur et al. 1998).

Economic opportunities have also created circular production structures, though, where products are collected at the end of their lives and reprocessed into secondary supply for product manufacturing. This value recovery from end-of-life products can be based on the recycling of materials, the reuse of components or the refurbishment or repair of the whole product (Thierry et al. 1995). Product take-back legislation and other public policy instruments driven by environmental concerns are increasing the significance and the uptake of these production

B. Fleischmann et al. (eds.), *Distribution Logistics*

practices (Davis 1996). An important type of circular production structure is the closed-loop supply chain for end-of-life products. Here end-of-life products are reprocessed and used as secondary input for the manufacturing of a product with original functionality (Guide & Van Wassenhove 2002). Some of the management challenges of closed-loop supply chains are very different from conventional supply chain issues (Guide 2000, Krikke et al. 2001b) and lead to whole new groups of mathematical models (Fleischmann et al. 1997). Many of these new models focus on specific operational issues like inventory control or recovery network design, though, and only recently have quantitative models for more integrated supply chain management issues been developed (Savaskan et al. 1999, Krikke et al. 2001a, Debo et al. 2002). The systems approach has yet to be fully applied to the management of closed-loop supply chains, though, and the following chapter is a step in this direction.

2 Constraints in Closed-loop Supply Chains

From a physical point of view a supply chain is a set of transformation processes which are linked to each other by material flows. The inputs of one process are the outputs of all its upstream processes. A constraint is created when not the entire output of a process can be used as input for the downstream process it is designated for. In such a case a fraction of the process output will be lost as waste. The useful output of a process, often called yield, should therefore be modelled either as a constant fraction of the input or, more realistically, as a non-linear function of input and other variables, which influence the outcome. We show in the remainder of this chapter that non-linear constraints can have a significant impact on the supply chain performance.

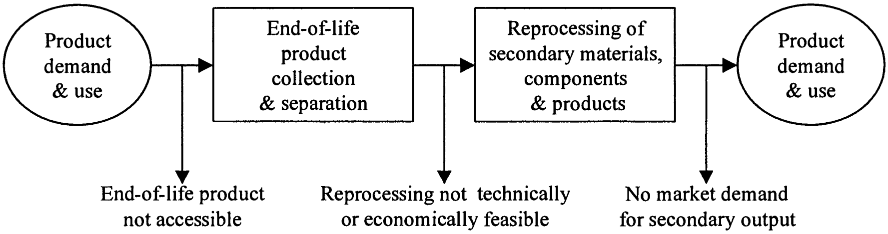


Fig. 1. The three types of constraints in closed-loop supply chains

In conventional forward supply chains upstream processes produce output which is customised for the downstream process since the final good is the ultimate purpose of the supply chain. Closed-loop supply chains rely on end-of life products as production inputs which are typically not designed for end-of-life value recovery. Closed-loop supply chains also have to be organised in order to cope with the uncertain timing of the return of end-of-life products and their uncertain content of recoverable value. For these reasons they are much more likely to be constrained than forward supply chains. Constraints in closed-loop supply chains can be divided into three different types (see Figure 1):

- Limited access to end-of-life products
- Limited feasibility of reprocessing
- Limited market demand for the secondary output

In the remainder of this paper we discuss the interaction between these three constraints for the case of reusing components from end-of-life products for product manufacturing. In the first part we explore how limited component durability impacts the feasibility of reprocessing. The second part investigates how finite product life cycles limit the market demand for secondary supply.

3 The Case of Closed-loop Component Reuse for Product Manufacturing

It is not unusual that components of durable goods last considerably beyond the use time of these products. Therefore, some companies practice or at least consider the systematic reuse of components from end-of-life products for the manufacturing of original equipment. Examples are microchips in computers and mobile phones, electric motors in copiers and power tools and various components in single use cameras (Keeble 1998, Clift & Wright 2000, Murray 1995, Klausner et al. 1998, Kodak 2001). When most of the product components are being reused so that the identity of the reprocessed end-of-life product is maintained, it is more common to call the process remanufacturing or refurbishment rather than component reuse (Thierry et al. 1995, Guide 2000). Our modelling results apply to any form of closed-loop component reuse for product manufacturing, and we therefore use these notions interchangeably. A different, important form of component reuse, which is not discussed in this paper, is the use of components from end-of-life products as spare and service parts.

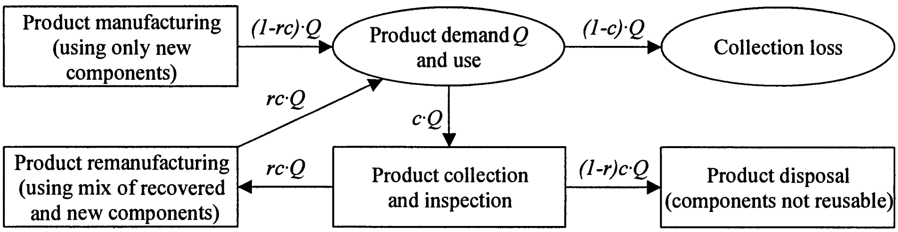


Fig. 2. Basic model of a production system with component reuse

To be able to quantify the effect of flow constraints we introduce a process and product flow model of a production and consumption system with product recovery and remanufacturing of a perfect substitute. The model distinguishes between four different groups of operations, i.e. manufacturing, collection, remanufacturing and disposal (detailed in Table 1), which have different types and quantities of product inflows and outflows as shown in Figure 2. The take-back scheme of the manufacturer collects a certain percentage c of all marketed products Q after they reach the end of their useful lives. In most cases the collection rate c will be smaller than one because some collection networks are not accessible to all owners of end-of-life products, some owners do not return the product but store it or dispose of it, and some end-of-life products are also collected by third parties. The collection process is followed by an inspection process, which may involve partial disassembly, and only a percentage r of all collected and inspected products can be remanufactured and remarketed. The remanufacturing yield r is limited by technical constraints, the most important of which will be discussed in the next section, and also dependent on sufficient market demand, which will also be the subject of this paper.

Table 1. Description of all processes shown in Figure 2

Process	Operations
Manufacturing	– Purchase of components and materials – Manufacture of in-house components – Assembly and testing of products – Distribution of the manufactured products
Collection & inspection	– Collection of end-of-life products – Inspection and testing of their durable components
Remanufacturing	– Disassembly of remanufacturable end-of-life products – Cleaning and repair of reusable components – Purchase or manufacture of worn-out components – Assembly and testing of remanufactured products – Distribution of remanufactured products
Disposal	– Disposal of non-remanufacturable end-of-life products
Collection loss	– These products are outside of the firm's responsibility

The other part of the collected and inspected end-of-life products cannot be remanufactured and creates disposal costs for the company, which might be somewhat reduced through materials or energy recovery. The disposal of uncollected end-of-life products has to be paid by the end user, which is the business reality for practically all goods at present. In Europe, North America and Japan governments are introducing or considering product take-back legislation (OECD 1996), though, which would considerably change the economics of product remanufacturing. Examples are the European ELV and WEEE directives on end-of-life vehicles and waste electric and electronic equipment. The implications of such legislation are not considered at present, but can be easily included in the model. The product between collection rate c and remanufacturing yield r, called remanufacturing rate rc, indicates the efficiency of product recovery management based on component reuse. The remanufacturing process brings the end-of-life product back to a like-new condition and the remanufactured product provides the customer with the same utility as the newly manufactured one. They are perfect substitutes on the market and the total demand Q can therefore be satisfied with any mix of manufactured and remanufactured products.

In this quasi-static description of the system the flow variables show no time dependency. For dynamic modelling purposes we can express the total demand as the time integral of the demand rate, $Q = \int_0^T q(t)dt$, T being the end of the planning horizon or the product life cycle. All other variables can then also be expressed as time dependent values and their overall values shown in Figure 2 are determined through integration over time. This paper is not concerned with the operational details of product recovery and remanufacturing. There is ample literature covering issues like reverse logistics, disassembly strategies, production planning and inventory control of remanufacturing systems. Good overviews are provided by (Fleischmann et al. 1997), (Guide et al. 1998) and (Guide 2000). It is also of no importance to our analysis whether the company itself or subcontracted third parties carry out the collection and remanufacturing activities (see e.g. Savaskan et al. 1999 and Guide & Van Wassenhove 2001) since we do not deal with supply chain coordination issues here. The critical observation for our analysis is that the processes detailed in Table 1 deal with different types and more importantly quantities of product flows.

3.1 The impact of limited component durability

As mentioned earlier not all of the collected products can be remanufactured. One important reason for this is the limited durability of the reusable components. In order to quantify this durability constraint to remanufacturing a suitable parameter is now introduced. It denotes the average number of times a component can be used for the same kind of product and is called average number of component lives n:

$$\text{Average number of component lives } n := \left\lfloor \frac{\text{average component life}}{\text{average product use}} \right\rfloor$$

The average component life is the mean of the distribution of component life times, which is a function of component design, product design and the characteristics of the product use and maintenance. The total lifetime of a component should include lifetime extensions via repair and refurbishment. This is done for example with electric motors in industrial products (Klausner et al. 1998), where the motor is disassembled, cleaned and the commutators are resurfaced. For our analysis we assume that a reusable component from an end-of-life product can be brought back to a like-new condition. Modelling deterioration (see e.g. Swan 1970) of the component is beyond the scope of this paper but would be a useful extension of the model. The average product use denotes the average total usage between product sale and end-of-life and is derived from a customer use distribution. Average component life and average product use are given in the quantity which best describes the use of the product (e.g. total use time, mileage, cycles, etc.). Table 2 lists some products and the average number of lives n of their reusable parts. If the variances of the component life and customer use distributions are large, the resulting distribution of the number of component lives will also have a substantial variance. In such a case the company will have to use large safety margins to reduce the statistical volatility and thus ensure reliability of the reused components.

Table 2. Average number of lives n for some product components

Product	Component	*n*	Source
Power tool	Electric motor	2-3	Klausner et al. 1998
Computer	Chip	4	Keeble 1998
Single use camera	Camera core	6	Kodak 1999
Glass bottle	Bottle	25	UBA 1996
Wooden pallet	Pallet	50	UBA 1996
Crates for bottles	Crate	120	UBA 1996

To understand how limited component durability affects our production and consumption system with product take-back and remanufacturing we follow a batch M of newly manufactured products through all stages of their n life cycles in Figure 3. At the bottom line we show the total product quantities each different process has to deal with. We assume for the moment that there is always enough market demand to absorb all collected end-of-life products that can be remanufactured. Therefore products leave the system only because they are either not collected at the end of their last use phase, or their reusable components have exhausted their maximum durability characterised by n. We also assume that all reusable components have the same number of lives n. This assumption is not necessary to obtain the basic results of this section but significantly simplifies the bookkeeping and notation in the mathematical formalism. A last assumption is that the collection rate c is the same for all collection cycles.

The results from Figure 3 can now be employed to quantify how limited component durability limits product remanufacturing. Since the remanufacturing rate rc is defined as the ratio of all remanufactured products over all sold products, and the remanufacturing yield r as the ratio of all remanufactured products over all collected end-of-life products, rate and yield cannot exceed the following values:

$$rc \le (rc)_n := \sum_{i=1}^{n-1} c^i \Big/ \sum_{i=0}^{n-1} c^i$$
$$r \le (r)_n := \sum_{i=1}^{n-1} c^i \Big/ \sum_{i=1}^{n} c^i = \sum_{i=1}^{n-1} c^i \Big/ \left(c \sum_{i=0}^{n-1} c^i \right) = (rc)_n / c$$

This result is an upper bound for all take-back and remanufacturing systems with unconstrained product demand and limited component durability. We will call $(rc)_n$ and $(r)_n$ the durability limits to the remanufacturing rate rc and the remanufacturing yield r. There is an elegant alternative to quantify these limits: From Figure 2 it follows that total demand Q is satisfied with $(1\text{-}rc)Q$ newly manufactured products and rcQ remanufactured ones, which already completed one or more cycles in the production system. The minimal amount of products that has to be newly manufactured is $(1\text{-}c)Q+c^n(1\text{-}rc)Q$, which means that $(1\text{-}rc)Q \geq (1\text{-}c)Q+c^n(1\text{-}rc)Q$. The first term, $(1\text{-}c)Q$, is to compensate for leakage through imperfect collection and the second, $c^n(1\text{-}rc)Q$, replaces the losses through limited durability, which are all the products that contain components returning for the n[th] time and have to be disposed of. Transformation of the inequation yields

$$rc \leq (rc)_n := \frac{c-c^n}{1-c^n} \quad \text{or } r \leq (r)_n := \frac{1-c^{n-1}}{1-c^n}.$$

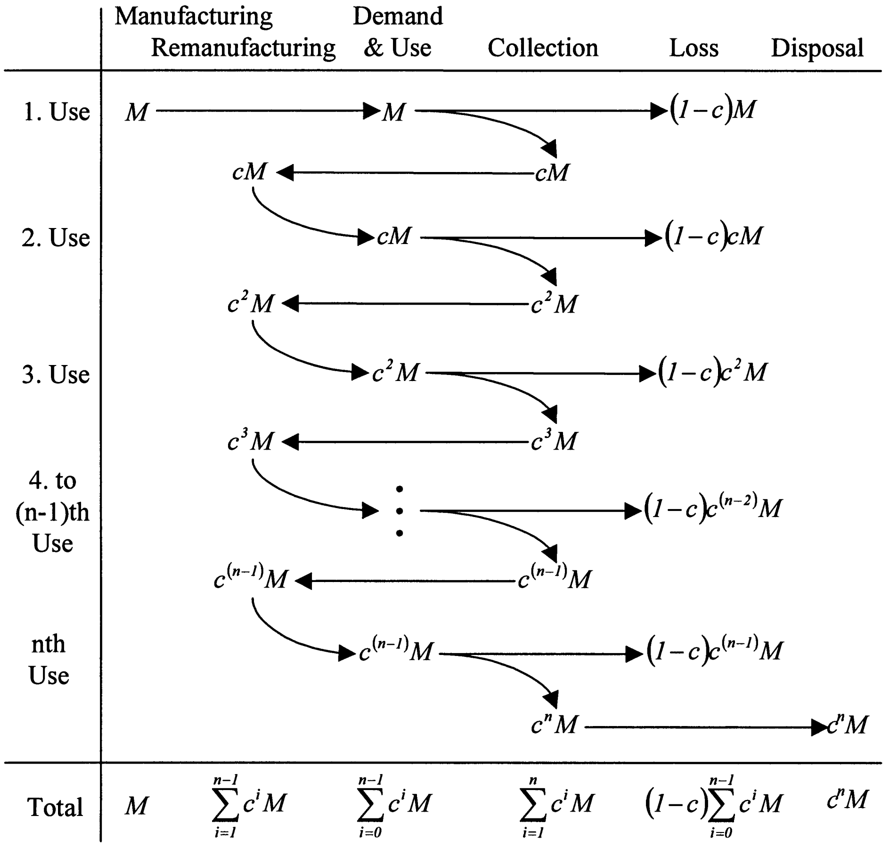


Fig. 3. Flow of M newly manufactured products through their n life cycles

It is easily shown that for $c \neq 1$ both expressions for the durability limit $(rc)_n$ to the remanufacturing rate rc are identical,

$$\frac{c-c^n}{1-c^n} = \frac{c-c^n}{1-c} \cdot \frac{1-c}{1-c^n} = \sum\nolimits_{i=1}^{n-1} c^i \Big/ \sum\nolimits_{i=0}^{n-1} c^i ,$$

and have the following properties:

$$\lim_{n\to\infty}(rc)_n = \lim_{n\to\infty}\left(\frac{c-c^n}{1-c^n}\right) = \mathrm{c} \qquad \text{(unlimited component durability)}$$

$$\lim_{c\to 1}(rc)_n = \lim_{c\to 1}\left(\frac{c-c^n}{1-c^n}\right) = \frac{n-1}{n} \qquad \text{(perfect collection)}$$

$$\frac{d}{dc}(rc)_n = \left(\sum_{i=1}^{n-1} ic^{i-1}\right) \Big/ \left(\sum_{i=0}^{n-1} c^i\right)^2 > 0$$

in particular $\frac{d}{dc}(rc)_n(c=0)=1$ and $\lim_{c\to 1}\left(\frac{d}{dc}(rc)_n\right) = \frac{n-1}{2n}$

Limited component durability turns the maximum possible remanufacturing yield into a function of the collection rate *c*. The numerical example from Table 3 (with *n=3* and *M=1000*) helps to develop intuition for this not entirely obvious result. As the collection rate increases more and more products whose components reach the end of their durability return to the company and lower the yield of the remanufacturing process. This leads to a maximal remanufacturing rate $(rc)_n$ which grows slower than the collection rate *c* does. Figure 4 and Table 4 shows that this effect is the more pronounced the higher the collection rate and the lower the component durability.

Table 3. Durability limit to *r* for *n=3, M=1000* and different collection rates *c*

Collection rate *c*	**0.1**	**0.5**	**0.9**
1. Use	1000	1000	1000
1. Collection and remanufacture	100	500	900
2. Collection and remanufacture	10	250	810
3. Collection and disposal	1	125	729
Total remanufactured products	110	750	1710
Total collected products	111	875	2439
Maximal remanufacturing yield $(r)_n$	**0.99**	**0.88**	**0.70**

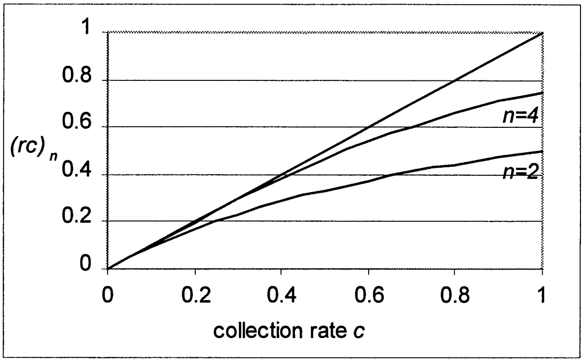


Fig. 4. Durability limit $(rc)_n$ as a function of the collection rate c

Table 4. Durability limit $(rc)_n$ as a function of the collection rate c

c	0.1	0.3	0.5	0.7	1
$(rc)_n$ for $n=2$	0.09	0.23	0.33	0.41	0.5
$(rc)_n$ for $n=4$	0.10	0.29	0.47	0.61	0.75

As an example, consider the electric motors, which are used in power tools and could be reused for power tool manufacturing (Klausner & Hendrickson 2000). A power tool manufacturer that considers setting up a take-back system and reusing components for product manufacturing has to trade off the costs of collection with the savings from component reuse. Once it is established that the savings from reusing one electric motor are larger than the cost of collecting one end-of-life power tool it seems that the company should aim for the highest possible collection rate. However, whereas the total collection cost increases with the collection rate c, the total cost savings from electric motor reuse increase instead with the maximal remanufacturing rate $(rc)_n$. Since the maximal remanufacturing rate grows slower than the collection rate it is possible that there is an optimal collection rate beyond which the net savings from power tool take-back and electric motor reuse start to decrease again. The reverse logistics of the take-back system therefore need to be coordinated with the durability of the reusable components and should not be set up independently.

The durability limit to the remanufacturing rate is a non-linear function not only of the collection rate c but also of the average number of component lives n. It can be seen in Figure 5 that the impact of the limited component durability is even more dramatic here since n enters the durability limit $(rc)_n$ as a power. Initially (n=1), an increase in durability drastically increases the maximal possible remanufacturing rate but the increase soon begins to level off and eventually reaches the collection rate as an asymptotic value. This effect is faster and more pronounced for lower collection rates.

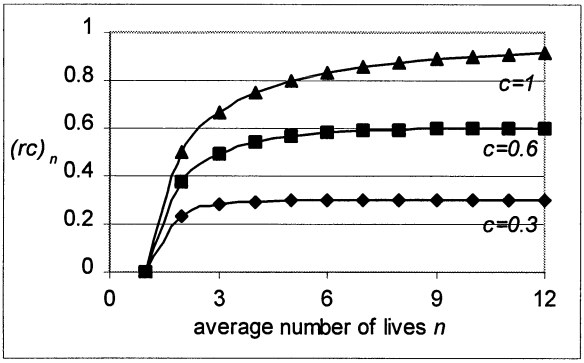


Fig. 5. Durability limit $(rc)_n$ as a function of the average number of lives n

That the marginal remanufacturing rate decreases drastically with increasing durability has very important managerial implications. For each collection rate there is a range of durability choices beyond which increasing component durability has virtually no impact on the remanufacturing rate. This means that the performance of the remanufacturing process is dependent on both the component durability and the performance of the collection process. For product take-back and remanufacturing to be successful all three aspects, product design, the collection process and technical feasibility of remanufacturing, have to be managed jointly. A last example will illustrate this important observation. The collection system of Kodak's single-use cameras yields a current collection rate of approximately 60% worldwide (Kodak 2001). The camera frame, metering system and flash circuit are designed to be used around six times ($n=6$) (Kodak 1999). As a result Kodak could satisfy a maximum of 58% of the market demand with cameras containing reused components. As discussed above, no matter how often the camera components could be reused, the maximal possible remanufacturing rate could never exceed the collection rate of 60%, which is only 2% more than the current figure. On the other hand, if n would be three instead of six the maximal remanufacturing rate would drop below 49%, so from a design perspective the durability choice n is consistent with the collection rate c.

3.2 The impact of finite product life cycles

In the previous section we assumed that there is enough market demand to absorb all collected end-of-life products that can be remanufactured. In reality the available amount of remanufacturable products may exceed their market demand, though, creating the third type of constraint. In the case of perfect substitutes the existence of the constraint cannot be due to low customer valuation but has a different reason, which we briefly outline first and then assess with a quantitative model.

The marketing concept of the product life cycle describes the sales of a product model from market introduction until market withdrawal (Mahajan et al. 1993).

Virtually all product models stop being produced after a certain amount of time and are therefore offered on the market for a limited time only, which means that they have finite product life cycles. How does this affect remanufacturing? A very important feature of production systems with product take-back and remanufacturing is the time lag between product sales and availability of end-of-life products. Products return at the end of their use phase, the length of which varies enormously from product to product. Due to the finite life cycle of product models some end-of-life products may only return when no demand is left for this particular product model. This can have a strong impact on product take-back and remanufacturing, which we now demonstrate and quantify.

The product life cycle describes the sales of a product model over time and is therefore formalized as a function $q(t)$, with $q > 0$ for $0 < t < T$ and $q = 0$ elsewhere, T being the length of the product life cycle (Mahajan et al. 1993). A certain percentage of the sold products will return to the company according to a residence time distribution $d(\Delta t)$, which denotes the frequency of all customers (or the likelihood of one customer) to return the product after time interval Δt. The returned products at time t are now

$$v(t) = \int_0^t q(s) \cdot d(t-s) ds ,$$

which can be easily integrated numerically and will generally result in a curve that, relative to $q(t)$, is shifted to the right by the mean of $d(\Delta t)$ and spread according to its variance and higher moments. For our purposes we require a closed form solution for $v(t)$ that is simple enough to allow for analytical treatment but still contains all the essential features of time dependent product sales and returns. The product life cycle is therefore modelled as an isosceles triangle and completely described by the length of the product life cycle T and the total demand $Q = \int_0^T q(t) dt$:

$$q(t) = \begin{cases} 4Q \cdot t / T^2 & 0 < t \leq T/2 \\ 4Q \cdot (T - t)/T^2 & T/2 < t < T \\ 0 & \text{elsewhere} \end{cases}$$

Using an isosceles triangle, i.e. a triangle with legs of equal length, allows us to model the product life cycle with two parameters only. Any shape more complex than an isosceles triangle would require additional parameters, which, in this first analysis, would not help in developing intuition. To further simplify the model we also assume that the residence time between the sale of a particular product and its return to the manufacturer is a constant value t_Δ for all products (i.e. the residence time distribution is a Dirac-like δ-distribution at t_Δ) and that the collection rate has a constant value c (i.e. the return probability is c and constant over time). The collected end-of-life products as a function of time are then:

$$v(t)=\begin{cases}4cQ\cdot(t-t_\Delta)/T^2 & t_\Delta < t \le t_\Delta + T/2\\ 4cQ\cdot(T+t_\Delta - t)/T^2 & t_\Delta + T/2 < t < t_\Delta + T\\ 0 & \text{elsewhere}\end{cases}$$

A complete description of product life cycle and product returns is now given by the four parameters Q,c,T and t_Δ. Using a modelling approach based on a minimal set of system parameters is the best way to reveal the principal interactions of the system. This initial analysis can then be used as the foundation for further research. The sales data for a photocopier model in Figure 6 also show that our assumptions may not be unreasonable. Typically, a product model is not withdrawn from the market because the type of product is no longer in demand but because it is replaced by new product generations, which result from competitive pressures and technological innovations. For competitive reasons companies increasingly shorten the time for new product development and the time between market introductions of consecutive product generations, thereby shortening the life cycles of the product models (Billington et al. 1998). The shortening of the product life cycle mainly reduces the maturity stage of stagnant sales, which turns the traditional trapezoidal shape more and more into a triangle.

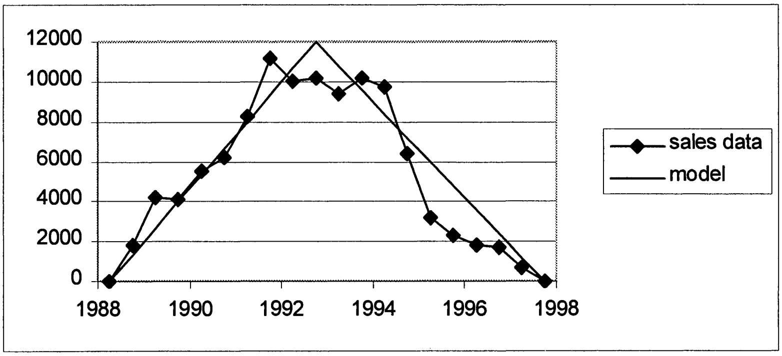


Fig. 6. Product life cycle (sales per year) of a photocopier model (Marx-Gomez & Rautenstrauch 1999)

Assuming that the durability of the components is infinite ($n = \infty$) the shaded area in Figure 7 represents all returned products that can be remanufactured and re-marketed. There is not enough market demand to accommodate the rest of the returned products, and they will have to be disposed of instead. It is important to be aware that the product life cycle $q(t)$ denotes the joint sales of all-new and remanufactured products. If a product is sold m times, it appears in Figure 7 as one sale of an all-new product and $(m-1)$ sales of remanufactured products (located in the shaded area). The product might then return again (for the m^{th} time) after the end of the product life cycle and has to be disposed of due to lack of market demand. The integral over the shaded area in Figure 7 divided by total demand Q gives an upper bound for the remanufacturing rate rc. In analogy to the durability limit it will be called remarketing limit to the remanufacturing rate and is calculated as

$$(rc)_{plc} = \int_{\Delta}^{T} |q(t) - v(t)|\, dt .$$

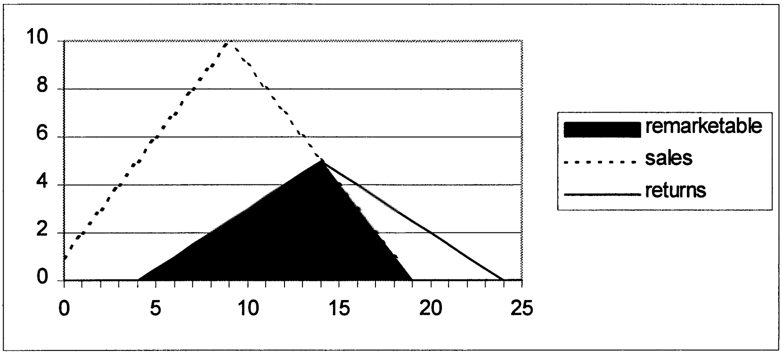


Fig. 7. Model of product life cycle and product returns

Using elementary geometry the closed form solution can be calculated as

$$rc \le (rc)_{plc} := \begin{cases} c - \dfrac{2c}{1-c}\tau^2 & \text{for } \tau \le \dfrac{1-c}{2} \\ \dfrac{2c}{1+c}(1-\tau)^2 & \text{for } \tau \ge \dfrac{1-c}{2} \end{cases}, \text{ where } \tau := \frac{t_\Delta}{T}.$$

The remarketing limit to the remanufacturing yield is calculated as $r_{plc} = (rc)_{plc} / c$. Figure 8 and 9 show that $(rc)_{plc}$ is a continuous function of the collection rate c and τ, which we will call the residence index. $(rc)_{plc}$ has the following properties:

$$(rc)_{plc}(c = 1) = (1-\tau)^2 \text{ and } (rc)_{plc}(\tau = 0) = c$$

$$\frac{d}{dc}(rc)_{plc} = \begin{cases} 1 - \dfrac{2\tau^2}{(1-c)^2} & \text{for } \tau \le \dfrac{1-c}{2} \\ 2\dfrac{(1-\tau)^2}{(1+c)^2} & \text{for } \tau \ge \dfrac{1-c}{2} \end{cases}$$

In particular, the boundary values of the first derivative of $(rc)_{plc}$ are

$$\frac{d}{dc}(rc)_{plc}(c = 0) = \begin{cases} 1 - 2\tau^2 & \tau \le 1/2 \\ 2(1-\tau)^2 & \tau \ge 1/2 \end{cases} \text{ and } \frac{d}{dc}(rc)_{plc}(c = 1) = \frac{1}{2}(1-\tau)^2 .$$

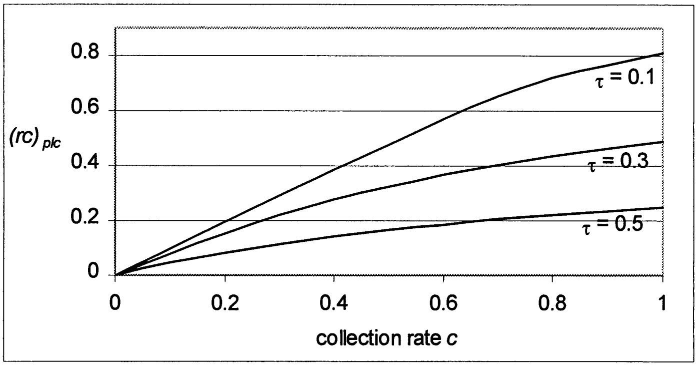


Fig. 8. Remarketing limit $(rc)_{plc}$ as a function of the collection rate c

Limited market demand due to the finite product life cycle turns the maximum possible remanufacturing rate into a non-linear function of the collection rate c. As the collection rate increases, proportionally more end-of-life products return outside the product life cycle than inside, which lowers the yield of the remanufacturing process. Like in the case of limited component durability this leads to a maximal remanufacturing rate $(rc)_{plc}$ which grows slower than the collection rate c does. This means that increasing the collection rate of the take-back system reduces the efficiency of the remanufacturing operations. Figure 8 shows that this effect is the more pronounced the higher the remarketing index τ. The performance of the remanufacturing process is dependent in a non-linear way on both the ratio between residence time and product life cycle and the performance of the collection process. In a production environment with product take-back and remanufacturing the reverse logistics of the return process should therefore be managed and monitored together with the characteristics of product sales and use. One assumption we made throughout the paper is for the collection rate to be constant over time. If the company attempts to influence the collection rate over time, e.g. with financial incentives like rebates on new sales for end-of-life trade-ins (Klausner & Hendrickson 2000), it would be beneficial to collect more products at the beginning of the life cycle and fewer at the end. Where take-back is offered as a general customer service (customers do not have to dispose of the product themselves) controlling the collection rate may prove to be more difficult. An alternative way to deal with the issue of limited product life cycles would be to reuse components across subsequent product generations. This approach requires appropriate product design and components with mature technology and low innovation rates.

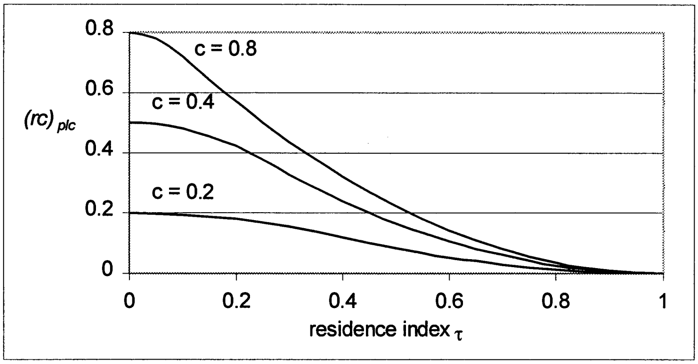


Fig. 9. Remarketing limit $(rc)_{plc}$ as function of the residence index τ

The remarketing limit decreases monotonically with increasing residence time and goes to zero as the residence time approaches the length of the product life cycle (see Figure 9). For a residence index equal or larger than one all returned products have to be disposed of without any value-added recovery through component reuse. This would typically be the case for many electric appliances and electronic goods, where current life cycles are often under one year (Billington et al. 1998). In the previous section we established that computer chips have an average number of lives of four and are therefore, from a durability point of view, a promising candidate for reuse. However, when most chips become available for reuse they are one or two generations old, and the computers they were built in are no longer on the market. The chips are either disposed of with the obsolete computers or sometimes reused in other low-tech applications, like electronic toys, for which, one could argue, they are highly overqualified.

4 Conclusions

In this paper we explore the impact of constraints on the performance of closed-loop supply chains based on component reuse. Typical constraints for component reuse in a closed-loop context are limited access to end-of-life products, limited feasibility of reprocessing and limited market demand for the final good containing reused components. We develop quantitative models for two important examples of such constraints, limited component durability and finite product life cycles. Limited component durability affects the technical feasibility of reuse, and the finite life cycle of product models impacts the market demand for the original product and its components. We show that both of these constraints introduce non-linearities into the flow rates of the production system. More specifically, the yield of the remanufacturing process turns into a function of the end-of-life product collection rate and other system parameters describing component durability, product use and the product life cycle.

This research can be continued in many directions. It would be interesting to investigate the impact of different collection systems and a collection rate that can be influenced by the manufacturer, e.g. through rebates and buy-back programs. The relationship between product design and feasibility of component reuse for product manufacturing should be explored further, in particular the possibilities to design for component reuse across product generations and types. We cannot generally assume that the remanufactured product is a perfect substitute to a all-new version. It would thus be useful to relax this assumption and the research of this paper with models of product competition.

Overall, our modelling efforts show that closed-loop supply chain models that fail to consider real-life constraints and the ways in which they interact might produce unrealistic results. Because of their non-linear interdependencies the end-of-life product collection, the feasibility of reprocessing and the market demand for the secondary output need to be assessed and managed in an integrated way. To advance this kind of systemic supply chain analysis it would be very useful to combine this type of operational analysis with an economic analysis to yield an operationally realistic and systemic economic assessment of closed-loop supply chains. Our operational modelling approach appears to be well suited to be combined with an economic model of closed-loop supply chains, and we will do so in our future research.

References

Billington, C. / Lee, H. L. / Tang, C. S. (1998): Successful Strategies for Product Rollovers, *Sloan Management Review*, Spring 1998, pp 23-30

Davis, J. B. (1996): Product Stewardship and the Coming Age of Takeback, Business and the Environment, Cutter Information Corp.

Debo, L. / Toktay, B. / Van Wassenhove, L. N. (2001): Market segmentation and production technology selection of remanufacturable products, INSEAD working paper 2001/47/TM/CIMSO 18

Clift, R. / Wright, L. (2000): Relationships between Environmental Impacts and Added Value along the Supply Chain, Technological Forecasting and Social Change 65, pp281-295

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

Guide, V. D. R. / Jayaraman, V. / Srivastava, R. / Benton, W. C. (1998): Supply Chain Management for Recoverable Manufacturing Systems, *Interfaces* 30(3), pp 125-42

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

Guide, V. D. R. / Van Wassenhove, L. N. (2001): Managing Product Returns for Remanufacturing, *Production and Operations Management* 10(2), pp 142-154

Guide, V. D. R. / Van Wassenhove, L. N. (2002): Closed-loop supply chains, Section 40 in: Ayres, R. U. / Ayres, W. L. A. (eds.): Handbook of Industrial Ecology, Edward Elgar, UK

Keeble, J. (1998): From hackers to knackers, supplement *online*, *The Guardian*, May 21 1998

Klausner, M. / Grimm, W. M. /, Hendrickson, C. (1998): Reuse of Electric Motors in Customer Products, *Journal of Industrial Ecology* 2(2), pp 89-102

Klausner, M. / Hendrickson, C. T. (2000): Reverse-Logistics Strategy for Product Take-Back, *Interfaces* 30(3), pp 156-165

Kodak (1999): 1998 Corporate Environment Annual Report http://www.kodak.com/US/en/corp/environment/

Kodak (2001): 2000 Health, Safety & Environment Annual Report http://www.kodak.com/US/en/corp/environment/

Krikke, H. / Bloemhof-Ruwaard, J. / Van Wassenhove, L. N. (2001a): Design of closed loop supply chains: a production and return network for refrigerators, Erasmus Research Institute of Management (ERIM) working paper, ERS-2001-45-LIS

Krikke, H. / Pappis, C. P. / Tsoulfas, G. T. / Bloemhof-Ruwaard, J. (2001b): Design principles for closed loop supply chains: optimising economic, logistic and environmental performance, Erasmus Research Institute of Management (ERIM) working paper, ERS-2001-62-LIS

Lee, H. / Billington, C. (1995): The evolution of supply chain management models and practice at Hewlett-Packard, *Interfaces* 25(5), pp 42-63, Sept/Oct 1995

Mahajan, V. / Muller, E. / Bass, F. M. (1993): New-Product Diffusion Models, Chapter 8 in: Eliashberg, J. / Lilien, G. L. (eds.): Handbooks in OR and MS, Vol 5, Elsevier Science Publishers

Marx-Goméz, J. / Rautenstrauch, C. (1999): Predicting the Return os Scrapped Products through Simulation - a Case Study, Proceedings, 2nd Int. Working Seminar on Re-use, Eindhoven, The Netherlands, March 1-3, 1999

Murray, F. E. S. (1995): Xerox: Design for the Environment, Harvard Business School Case 9-749-022

OECD (1996): Extended Producer Responsibility in the OECD Area, Legal and Administrative Approaches [...], OECD Environment Monographs No 114, OCDE/GD(96)48

Savaskan, R. C. / Bhattacharya, S. / Van Wassenhove, L. N. (1999): Channel choice and coordination in a remanufacturing environment, INSEAD working paper 99/14/TM

Socolow, R. / Andrews, C. / Berkhout, F. / Thomas, V. (1994): Industrial Ecology and Global Change, *Cambridge University Press*, Cambridge

Swan, P. L. (1970): Durability of Consumption Goods, *American Economic Review* 60, pp 884-894

Tayur, S. / Ganeshan, R. / Magazine, M: (1998): Quantitative Models for Supply Chain Management, *International Series in Operations Research and Management Science* Vol 17, Kluwer, Holland

Thierry, M. C. / Salomon, M. / Van Nunen, J. / Van Wassenhove, L. N. (1995): Strategic Production and Operations Management Issues in Product Recovery Management, *California Management Review* 37(2), pp 114-135

UBA (Umweltbundesamt) (1996): Life-cycle assessment for drinks packaging systems, Texte 19/96, Umweltbundesamt, Berlin

Von Bertalanffy, L. (1976): General systems Theory: Foundations, Development, Application, 2nd edition, George Braziller, New York

E-business and Circular Supply Chains: Increased Business Opportunities by IT-based Customer Oriented Return-flow Management

Harold Krikke[1], Jo A. E. E. van Nunen[2], Rob Zuidwijk[2] and Roelof Kuik[2]

[1]CentER Applied Research, affiliated to Tilburg University, Tilburg, The Netherlands
[2]Erasmus University Rotterdam, Rotterdam School of Management/Fac. Bedrijfskunde, Rotterdam, The Netherlands

Abstract. This paper deals with the application of information technology (IT) in circular supply chains (CSCs). We consider information on the installed base critical, and present an illustrative example. Next we discuss a framework of different kinds of value contained in a return, and IT-applications useful in supporting its recovery or neutralisation in case of negative externalities. Also we show which kind of CSC is needed for which kind of return. We illustrate our work by three real life case studies.

Key words: reverse logistics, supply chain management, circular supply chains, product life cycle management, value creation, e-business/IT

1 Introduction

Traditional wisdom holds that return flows are something compulsory and obnoxious. (Stock, 2001) conclude that "the state of development of Reverse Logistics is analogous to that of inbound logistics of 10-20 years ago". However, the importance of the field is gaining recognition, already acknowledged by(Rogers and Tibben-Lembke, 1998) state that "while much of the world does not yet care much about the reverse flow of products, many firms have begun to realise that reverse logistics is an important and often strategic part of their business". Also, pioneering business companies have found opportunities to gain a competitive advantage through reuse and recycling. Let us illustrate this by a number of examples.

(Giuntini and Andel, 1995) report an example of an airline manufacturer where, due to competition for scarce manufacturing capacity, new production was given priority over out-of-production spare parts. The company created a repair and remanufacturing scheme for impaired components, avoiding drastic cost increases for spare parts, improving availability from 81% to 93% and also preventing 3rd parties from taking over their spare parts business.

(Marien, 1998) reports on a paint case, regarding of reverse logistics management, based on the concept of source reduction, as well as identifying

B. Fleischmann et al. (eds.), *Distribution Logistics*

various reverse logistics strategies for leftovers and residuals. Business benefits include saving on procurement and (pre-) manufacturing costs of materials and components, saving on disposal costs, improving disposal services to the customer and retailer, attracting 'green' consumers, being pro-active to environmental legislation and showing social responsibility.

(Caldwell, 1999) presents a case study on Estee Lauder regarding commercial returns. Their reverse logistics program comprehended streamlining of reverse logistics processes including returns authorisation, information systems and developing reuse markets amongst which company stores.

From extensive case study research, for overviews see (Bloemhof et.al, 1999), (Fleischmann et.al, 2000), and (Guide and Van Wassenhove 2002), it can be concluded that successful reverse logistics programs have taken an integrated approach: hence integration of forward and reverse chains and by full Product Life Cycle management (PLCm); see (Krikke et.al, 2004). PLCm is the process of optimising service, cost and environmental performance of a product over its full life cycle. Key issues include product design for recovery, re-engineering, product data management, installed base management and evaluating (end-of) life scenarios.

Therefore we prefer to speak of circular supply chains or closed loop supply chains rather than reverse logistics, because it covers the combined forward and reverse supply chain. It is important to distinguish between reuse in the original or similar chain or alternative supply chains. In this paper we use the term closed loop for reuse in the original or a similar supply chain and open-loop in alternative applications. The collective term is circular supply chain, defined as follows.

Circular supply chains are the set of integrated business processes that improve the availability and effective lifetime of installed bases, as well as the responsible disposition of discarded returns. This means that the forward production-distribution chain is extended by:

- service/maintenance/extended support functions for the installed base
- a reverse supply chain
- remanufacturing, new production with partly 'recycled' content.

The installed base is defined as the total number of placed units of a particular product in the entire primary market or a market segment. Installed base management concerns the care of these products during operations. In its traditional definition it comprehends replenishment, maintenance repairs, overhaul, spare parts management, and system upgrades. Installed base support is mostly looked at from a technical point of view and focused on capital goods. However, the installed base can also be monitored on more commercial parameters in terms of extended support. For example, when a customer with a high volume copier only makes a few copies a week, there is a clear case for early take-back and exchange. This approach might also apply to consumer goods, which we will consider in this paper too. For example, one may offer additional support services such as pro-active repair for washing machines. Product leasing is becoming common in consumer markets as well, and the importance of after-sales is of increasing importance, e.g. in the printing market through toner cartridges refill. New business models are being developed today , in which circular supply chain seem to fit well. For example, some car insurance companies offer green policies, through which a damaged car is repaired with parts cannibalised from dismantled wrecks. Xerox leases a green line of copiers, providing a lower cost per copy. The trend towards selling the function of products instead of the physical products themselves may create incentives for customers to accept recovered parts and products as equivalent to new ones.

In forward supply chains, the application of IT has undergone a strong impulse over the last decades. E-business comprehends doing business electronically by using IT such as the internet, EDI, mobile phones, satellites and so on. Major components of e-business are e-commerce (sales), Customer Relation Management and fulfilment. It appears that little research has been carried out on the potential of e-business for circular supply chains.

This paper addresses the question how to use IT/e-business to create value through CSCs by improving efficiency, customer service and environmental impact We will focus on the return management aspects, but as will become clear this can not be seen independent from the other elements of CSCs. We address the following research questions:

- Which value is contained in different return types?
- How is this value related to characteristics of the CSCs?
- Which information is needed to improve the CSC?
- Which forms of IT/e-business can be helpful?
- What is known from business and what can we learn?

The paper is set up as follows. In Section 2 we develop a basic framework on the value of returns and the kind of CSCs needed. In Section 3, we describe IT that may support in obtaining this information and in Section 4 we describe some practical cases and draw lessons from this. In Section 5, we elaborate on the use of installed base information for product maintenance. In Section 6, we draw conclusions.

2 A Basic Framework

The first issue we address is which types of returns can be distinguished and what are the characteristics of their CSCs. Excellent studies on reverse supply chains' typologies have been presented by Fleischmann et al. (2000), Bloemhof et al. (1999) and Guide and Van Wassenhove (2002) (summarised in the Appendix). Using the principle of reflective research cycle, described by Van Aken (1994) and applied to Supply Chain Management by e.g. Van der Vorst (2000), we derive that there are three basic kinds of value contained in returns:

- Negative value of possible externalities. This often refers to impact on the environment and related disposal cost to be paid by taxpayer money. But it can also involve controlled destruction by the Original Equipment Manufacturer (OEM) to avoid third parties taking over the repair business. At all times one wishes to avoid or minimise damage.

- Positive intrinsic value. This positive value of a return may lies in invested resources. The latter of course refers to materials and energy but also labour.
- Positive or negative time based value. This value relates to both previous ones, but deteriorates quickly over time. Positive time based value is most likely customer or market driven. Examples include seasonal or fashion products. An example of negative time based value is nuclear waste.

The negative or positive value contained in the returned item must be neutralised or regained respectively. It depends on many factors which processing option (disposal, open loop recovery, closed loop recovery) is most feasible. Note that a return may be in different classes regarding market, environmental and cost value, it may have positive economical value and also negative environmental externalities. Dealing with conflicting drivers is one of the most challenging aspects of CSCs.

The next issue is: what are the requirements for CSCs based on the value classification? In the appendix, we describe a typology of circular supply chains per typical return defined. We aggregate this on a higher level of abstraction and relate the CSC characteristics directly to the value contained in the return in Table 1, in a manner similar to (Fisher, 1997). Fisher presents a framework for forward supply chains, in which the type of supply chain (responsive or efficient) depends on the kind of product (innovative versus functional). Note that the characteristics of the CSCs are matched with returns and have nothing to do with the original product.

Table 1. Returns and circular supply chains similar to Fisher

	Responsive CSC	Efficient CSC	Control CSC
Returns Time Based Value	FIT		
Returns Intrinsic Value		FIT	
Returns Negative Externalities			FIT

A control CSC focuses on neutralisation of negative externalities or in short damage control. It needs to be secure, ensuring a responsible processing of the returns. These chains may involve some bureaucracy, and are often slow moving. The returns are pushed into the channel, and finding markets is often difficult. Careful registration of input and output, is needed to verify mass balances or to prevent leaks in the system.

An efficient CSC deals with valuable items in stable markets. It is able to separate valuable parts of the returns from the non-valuables at an early stage. The non-valuable fluff may be transferred to a control CSC. There is still a market, and the CSC must be able to balance supply push and demand pull well in a time-phased manner.

A responsive CSC acts quickly because the value of returns is running out (of time) or damage to the environment is getting worse by the minute. These CSCs are demand driven, but collection (acquisition) may be hard.

CSCs need information for planning and control. Exchange of consistent, reliable data makes that for individual parties in the reverse chain exogenous variables turn endogenous.

In Section 4, we will describe how the cases fit in our framework of Table 1 and how IT-applications can improve the CSC involved. We aim to supplement the framework of Table 1 by indicating which parameters are relevant and hence which information is of value. First we first discuss a range of technologies in Section 3.

3 Information Technology Applications

Information technology pervades logistics management and will continue to do so in the near future. One next major step is likely to be the introduction of smart cards and smart labels that will enable the identification at a distance of individual products through the use of radio frequency. Applications of this technology are emerging. For example, smart labels are being used to improve product rotation of perishable goods, to track equipment that is for rent and to monitor in-process parts and finished goods (computer parts and cars).

Data storage capacity is nowadays sufficient to accommodate the information needs of multiple actors in a supply chain. The technology has particular potential for products that are (i) time-sensitive, (ii) unique (in their characteristics: qualities, routing, use) and (iii) valuable. Many products in reverse logistics qualify on each of these attributes as candidates for the application of smart technology.

3.1 Product data management and data logging

Product Data Management (PDM) in general serves to maintain accurate data on complex products (many parts, variants, alternatives), record maintenance changes on a product during its lifecycle and disseminate product data at an intra-organisational or inter-organisational level. PDM improves the quality of data and reduces labour intensity, since it reduces the amount of manual data transfer of information in the chain. Whereas PDM systems capture information provided by communicating partners, they may not record information on products' condition and configuration that result from the operations of the product.

Data-logging automates the registration of important use related variables. Examples include peak loads, total running hours, average temperature. (Klausner et.al, 1998) report on a so called green port implemented in Bosch power tools for end-of-life optimisation. Based on the monitoring of a few parameters during the products life time, a DSS is capable of classifying returned cores into 'reusable' and 'scrap'. Substitutes may involve the use of scan units. A scan unit is able to make an X-ray of returned goods and make a reliable estimate of product parameters otherwise monitored by PDM.

The use of recycling passports for consumer electronics is imposed by EU legislation. The internet is used to exchange the passport between various members of the CSC (Spengler, 2002). The recycling passport enables to identify hazardous materials at an early stage but also to separate material fractions in a better way. Many material fractions in electronics need to be separated perfectly otherwise they are worthless and might as well be dumped. Reliable identification helps to obtain pure fractions.

3.2 Remote condition monitoring and diagnostics (smart products)

New communications protocols can be combined with embedded control systems to monitor product data dynamically and transmit the data logged upon some event. Based on this proactive action can be undertaken. Examples include early detection of technical failures, optimisation of fuel consumption and automated replenishment. Often condition monitoring is installed for maintenance purposes. But once installed it also provides the means for remote (on line) monitoring of critical parameters in the operations phase of the product to be passed on to the recovery stage. Disposition decisions can now be made at an early stage without physical inspection or disassembly. It can also be used for pricing mechanisms in e-market places. Moreover, re-engineering by feed back of return info to product design and production stages results in product quality improvements.

Monitoring is one thing, diagnoses is another. Diagnosis is complex because often it is not a single (defect) component that causes the trouble, but an unexpected interaction of components in the system. So the trick is not only to detect the direct cause of failures, but in particular the chain of reactions following it. It is essential to distinguish which product parameters to be monitored and their 'alarm levels', but also to limit the number of parameters to be monitored. Fuzzy logic and neural networks provide interesting opportunities. The use of complicated techniques to deal with complicated issues is tempting, but certainly requires justification by means of examples that show successes in practice. There is some reason to be careful mentioning the use of these "general purpose" techniques.

3.3 Tracking and tracing

In order to ease collection, it is important to know the condition, but also the location of the returns in advance. Many companies are implementing chips in containers and pallets, companies are able to track and trace them for logistics purposes. The capability of linking with geographic information systems, often done for other reasons, increases opportunities to e.g. follow containers online and plan their return.

3.4 E-marketplaces

E-commerce is often also seen as a cause of returns. However, it may also be used in prevention. Dell Computer's small- and home-business sales division found that returns for web-ordered PC's were lower than PCs ordered by phone, thanks to a configuration feature that DELL added to its site in 1996 (Caldwell, 1999). Most of all, the internet serves as a means to consolidate (geographically) fragmented markets for product acquisition and re-sales. Regarding new e-business models one can distinguish three types. (Kokkinaki, et. al, 2001).

Returns aggregators bring together suppliers and customers, automates the procurement of returns and creates value through consolidation: high throughput and minimal transaction costs. Many returns aggregators follow subcontracting third parties to do some or all of the described logistic functions.

Speciality locators are vertical portals, which focus on niche markets for highly specialized used parts or products, such as authentic antiques, exact replicas parts or equipment in historic restoration projects or the maintenance process for vehicles and industrial equipment. Their services include training, frameworks for catalogue search, selection and configuration, financing and technical support. Speciality locators are mostly used for information dissemination and address the marketing aspect of electronic commerce.

The integrated solutions providers go a step beyond facilitating and matching demand and supply of returns. The model for Integrated Solution Providers does not view e-commerce as a migration of existing practices and services but as a new tool to restructure a business activity and offer new services. Furthermore, they actually become the owners of the returns instead of implementing a brokering mechanism as the previous two models. By definition, each integrated solution provider focuses on the reverse logistics network in an industry/sector. More information can be found in (Krikke et.al, 2003).

3.5 E-ERP

Existing ERP-systems are not geared for recovery. An extensive description can be found in (Krikke, et.al, 2003). Here we briefly discuss how manufacturing resource planning (MRP) and warehouse management systems (WMS) ideally should support CSCs.

On the recovery side alternative bills of materials (BOMs) for one product are needed to represent alternative routings in the MRP-system depending on the recovery option chosen. On the return side, various routings can also be represented by alternative disassembly BOMs. With regards to inventory control, cores returned must be registered distinguishing between different types (statuses) and owners of returns inventory. Note that cores and components disassembled can be found in different states e.g., "to be dismantled", 'scrap, 'repairable' or 'reusable'. Moreover, identification of identical components retrievable from various cores at different indenture levels must be supported. Some components might not be identical but fulfil the same function. Often, specific components can be obtained by cannibalising cores or modules, and the availability of these components must be visible in the inventory control system, not only must the modules be registered as a whole, but also an explosion into components must be visualized. Forecasting can be improved by adding a line with a returns forecast to the Master Production Plan. This works best in the so called dual bill method. At the level of components, one would like a re-planning functionality after core disassembly, since actual quality of the components is known then.

A warehouse usually has to deal with three kinds of returns: retailer returns, supplier refusals and packing returns. Returns cause additional, labour intensive processes in the warehouses, which constitute around 6% of total logistics costs. The WMS gives administrative support primarily of the return process itself. Ideally, it determines next steps to be made further down the channel (recovery options) and communicates this information to the players involved, including the potential re-users.

4 Cases

We describe here two cases, namely OMRON and Philips, in which e-business concepts are introduced for the returns handling. We will first describe both cases in 4.1 and 4.2 and subsequently reflect them to our framework in Table 1 in 4.3. Also, we will discuss the three major e-business components e-commerce, Customer Relation Management and fulfilment.

4.1 OMRON

Headquartered in Japan, OMRON is a global corporation manufacturing products in 30 factories. OMRON has 71 subsidiaries in 35 countries. The company supplies products and services in the fields of industrial automation, automating services, medical health care and information processing. Typical products are card readers, vision systems, switches, sensors, counters, controllers, relays and connectors.

With headquarters in the Netherlands, OMRON Europe BV oversees sales, distribution and support activities in Europe, the Middle East and Africa. OMRON Europe BV comprises more than 20 fully owned subsidiary companies, including 18 national sales companies in virtually all of the European countries, and includes a European Logistics Centre (ELC) and a European Repair Centre (ERC). Repairs of products are a major component of OMRON's after sales service. Prior to 1999, the forward and backward repair flows in Europe followed the logic presented in Figure 1.

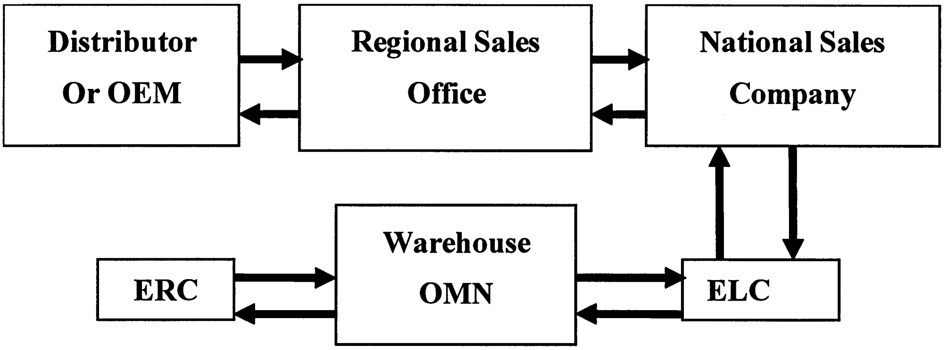


Fig. 1. The 'old' logistics network of OMRON

In Figure 2 OMN stands for the manufacturing plant OMRON Manufacturing Netherlands. This is a facility that is collocated with the European Logistics Centre and the European Repair Centre in Den Bosch. The return and repair process proceeds as follows. The user of a technical system informs his supplier when a failure occurs. The supplier, an original equipment manufacturer (OEM) or distributor, investigates the reported failure. If an OMRON product causes the failure then OMRON's regional sales office or national sales company is contacted. If it is decided that the failed products needs to be repaired then the product follows the chain of facilities as depicted in Figure 1. A temporary device often replaces the defective part. After repair the product is returned along the chain depicted back from the ERC to the OEM. The engineer then replaces the temporary spare with the repaired part. The customer is billed based on the details of the repair that has been made.

The process and its performance produced several problems and difficulties: long lead times (3 weeks – 3 months), the double substitution a faulty parts induced productivity problems at users, and the whole process needed repairs to be traced back to the customers/applications over a long period. Two options to improve were identified:

- Option 1: (i) reduce the lead time of all processes involved and (ii) optimize inventory control
- Option 2: (i) reformulate goals and (ii) change the structure of the processes

OMRON has the strategy to do repairs at the component level (resistors, processors) in stead of simply replacing modular subassemblies, for example circuit boards. This requires expertise and specialized equipment and creates the need for scale in performing repairs. So repairs should remain a centralized activity.

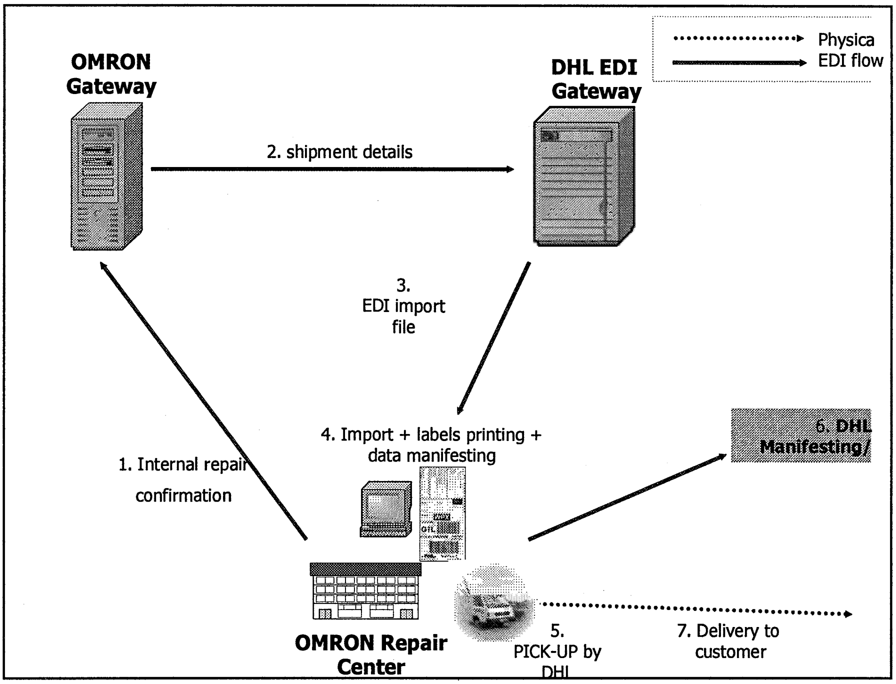


Fig. 2. Leg 2 of re-designed logistics at OMRON

In view of this, OMRON followed the second option. Together with the logistics service provider DHL, OMRON created from a multi-leg process a new 2-leg process: one leg going from the customer to the repair centre and one leg the other way round. Figure 2 illustrates the second leg going from the repair centre to the customer.

The most salient features of the new structure are

1. transportation from the ERC to the customer is direct, there are no intermediary facilities that consolidate transportation, the transportation is done by the service provider DHL.
2. the main point of contact between the OMRON organization and the service provider DHL that organizes the transportation is at the European level where both organizations have gateways for handling information transfer.

To support this structure the subsequent measures were introduced additionally. The objective was to shorten lead times to such an extent that engineers would no longer need to install a spare for a defective component. As a further administrative simplification customers were to be charged a fixed fee for repairs irrespective of the repair details.

In the new structure, fast and reliable transportation carried out at a premium substitutes for less expensive and slow transportation. However on the whole cost benefits accrue from the shortening and simplification of the repair process. Also the service level is improved.

Customer benefits:

- clarity on repair costs, ('flat fees')
- less shut downs due to replacement components,
- less overall costs

OMRON benefits:

- lead time reduction to approximately 5 days
- a decrease in product replacement activities by 90 %
- a decrease in repair engineering time by 30%
- a reduction of after sales costs by 30%.

Of course, the success of the system requires state of the art information technology for tight monitoring of the transportation and repair processes. Note that the system builds on systems that are in use at the two parties involved: the design is such that the required interaction between the systems OMRON and DHL for operating it is limited. The OMRON gateway continues to maintain OMRON's own data and the DHL gateway provides all the required information for the DHL organization to steer pickup and delivery. Information needed by DHL on the customer base of OMRON is transferred to the DHL organization at the level of the two European gateways. More details can be found in (Kuik et.al, 2004).

4.2 PHILIPS

Philips DAP (Domestic Appliances and Personal Care) is the division of Royal Philips Electronics in The Netherlands that focuses on the manufacture and supply of personal and domestics appliances such as electric shavers, beard trimmers, hair clippers, female epilators and shavers, hair, skin and body care products, tanning products, electric toothbrushes and water jets, kitchen appliances, irons and vacuum cleaners. The total product line comprises some 600 different products. Essentially the markets for the products used to be approached through national sales organizations (NSOs) with product supply occurring through a national distribution centre.

In the Netherlands and Germany the supply network of Philips DAP looked prior to 1998 as depicted in the following figure. In Figure 3 solid arrows represent regular sales flows and dotted arrows represent return flows.

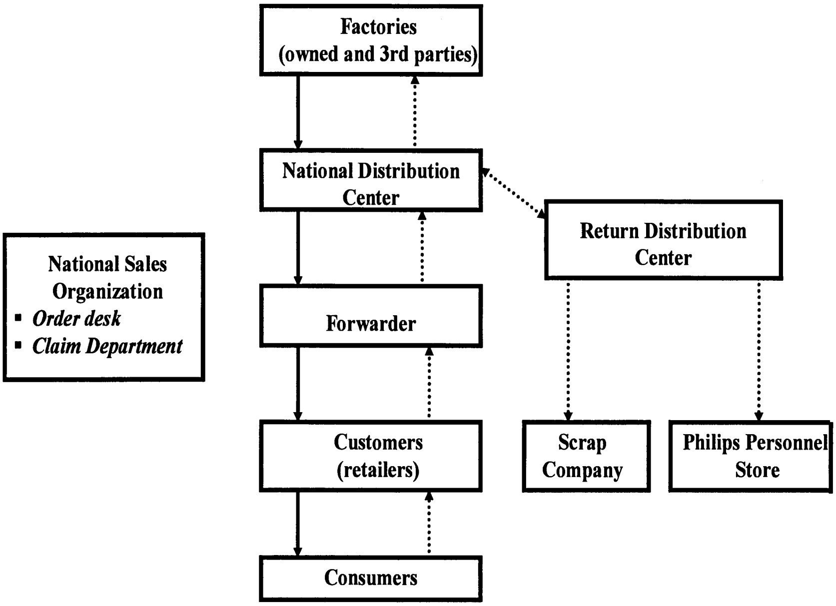


Fig. 3. Logistics network for Philips DAP

In 1998 Philips DAP worried increasingly about it's handling of returned products. At the same time it was felt that customer trends were such that more and more of the future sales volume would be realized during promotions. So there arose increasing pressure on the sales department to push product down the channels to support such promotional activities, which in turn was to lead to larger volumes of commercial returns. This concern was further compounded by an urge to keep up or rather forerun the need to boost the satisfaction of retailers/customers while reducing costs and stock levels by international consolidation throughout the supply chain.

The project carried out was to address the problems and to find solutions that would organically include handling of returns in logistics, operations and management of customer relationships. For the region The Netherlands, Belgium, Germany, and Luxemburg, which served as a pilot project, two major lines along which to change the supply chain were identified as:

- divide activities and responsibilities principally along the characteristics of front stage versus backstage in stead of division along national borders
- reduce the number of stock locations Europe-wide from 16 national warehouses to 5-6 regional distribution centres.

Of course both lines needed to be supported by information technology infrastructure such as document management systems to render the ensuing management at a distance viable.

Three main types of returns occurred: defective products, returns due to administrative errors, and commercial returns Return due to errors and due to defects are usually accompanied by complaints. So complaints roughly come in two types: complaints of an administrative nature and those with a nature of physical distribution (damaged products). There exists a European call centre that takes complaints. Customer service (a part of an NSO) deals with the first type of complaints and the Complaints Department (part of the DAP Service Centre) to be located near the Regional Distribution Centre is to deal with the second category. (The latter may vary slightly according to country). Commercial returns are potentially re-sold at other markets then they have been returned from or are shipped to the Phillips personnel stores. Depending on value, defective products are scrapped or repaired. Higher valued products, among which 3-head shavers, facial cleaners, espresso machines, vacuum cleaners, and air regulators, are eligible for repair. Customers are credited or receive replacement for defective products. Figure 4 illustrates the situation with respect to complaints and returns handling.

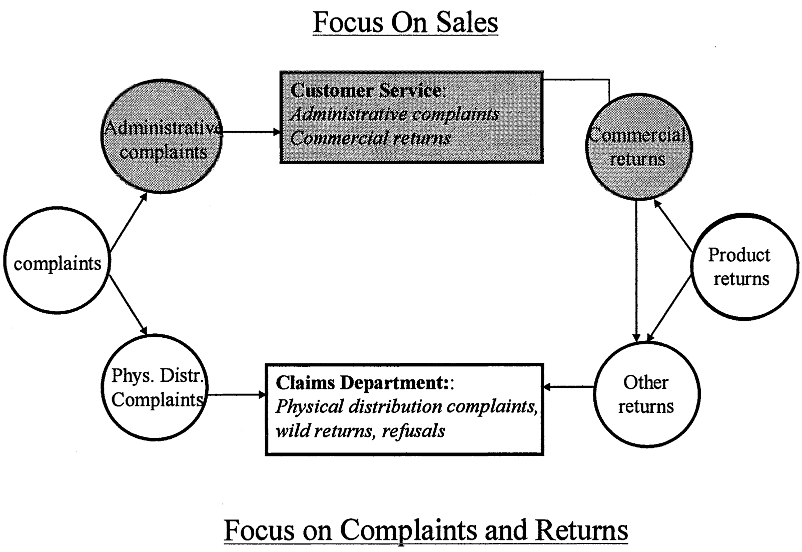


Fig. 4. Complaints and returns handling at Philips DAP

Introduction of the centralized Complaints Department (close to the regional distribution centre) may yield two advantages: efficiency and resource sharing.

The features for handling the return flows that stand out in the solution followed by Philips DAP are that (i) no separate logistics infrastructure is created for returns (although dedicated departments are created), (ii) returns are integrated in the forward flows and (iii) following the control of the forward flow, returns are handled centrally. More details can be found in (Van Galen Last,1998)

4.3 Discussion

Considering the value model presented in Section 2 of this paper the two cases presented clearly fall into different categories.

In the OMRON cases one deals with returns for which the handling has enormous time-based value for the customer. The value for OMRON's customers and OMRON itself is so high that it outweighs the extra costs incurred from utilizing fast transportation that to a large extent is independent of the forward flows. Here the time-based value is higher than the intrinsic value and this warrants a dedicated highly receptive channel for returns handling. The time value of (repaired) returns in the OMRON case warrants fast transportation

In the Philips case the intrinsic value of the returns is a dominant consideration. Hence, as the value model predicts, efficiency is a main concern for management. This manifests itself in the fact that forward and return flows share the major part of the logistics facilities.

Let us now discuss the three main elements of E-business per case: E-commerce, CMR and fulfilment. IT solutions are sought to deal with the distributed management involved, depending on the types of return. DHL, who besides the transportation supplies information and tracking and tracing services, here acts as an integrated solutions provider, be it that DHL does not assume ownership of the products. In fact a major part of the effort in implementation went into setting up the information structure within which OMRON and DHL communicate at the European level.

The urge to reduce the number of complaints and the need to closely monitor the processing of these complaints (both to be classified as negative externalities) play a major part at Philips DAP in their management of returns.

Practical lessons learned

- In general, one sees that IT solutions are always part of a bigger scenario, involving the facility and organization structure such as redesign of the network, reduction of returns, complaint handling and so on.
- Also, the cases do not solely and fully rely on managing the installed base, but also deal with the return channel itself.
- For returns with time based value the installed base matters in terms of information needed, for intrinsic returns this is less the case. In the OMRON case the gateways link the installed base with the repair process.
- Efficient CSCs seem to be have integrated with the forward and reverse channels to create economies of scale whereas responsive CSCs require separate return channels.

Modelling lessons learned

Returns can be classified in different manners:

- on stage on the product life cycle: commercial, end of use, end of life etc.
- on the driver for the customer: deposit, failure, legislation etc.
- on remaining value, as we did in table 1

The Philips case shows that the first two criteria are confusing. Being marketing related returns, with no reuse market they fit the definition of 'commercial' return and 'end of life' at the same time. However, the value (intrinsic) clearly maps which kind of return it is and hence an efficient CSC is appropriate.

5 An Illustrative Example: Value of using Information in Operations

The cases described in section 4 are very interesting qualitative examples, but there appears to be a lack of quantitative, OR-based studies to validate the implications of our studies. Therefore we present here a theoretical example.

In this section, we assess the value of using information on the installed base in a quantitative manner. The use of information enables more efficient use of resources and henceforth creates value. We discuss the situation in which products positioned at customers require maintenance and occasional repair, in short: service. It is assumed that the status of products (i.e. its technical condition) can be measured providing information on whether maintenance or repair is required within a certain time interval, in particular to prevent emergency calls for repair. We focus on the routing of technicians that service products at customer sites.

Traditionally, (preventive) maintenance is done on the products on a regular basis, while repair is performed at customer calling, in most cases on an emergency basis. The status of the machines is measured only when a technician is at the customer site. This traditional approach results in fixed routing of technicians to customers for routine maintenance with additional emergency repair calls. In general, before the actual visit, the status of the products on site is not known.

In case of remote and continuous status monitoring of the products, maintenance and/or repair decisions are based on direct pro-active assessment of the status of the products, where customer interference is not required. Routing of technicians based on actual status information with the appropriate priority and service parts becomes feasible. There are other efficiency gains. First of all, a priori information on the installed product may reduce the service time on that product. Secondly, reduced service and travel times may increase the service capacity of a technician per day. Thirdly, appropriate supply of service parts and better preparation may result in other savings than service labour time, such as lower service part inventory in service vehicles, etc. In principle, the installed base will be characterized by customer locations and status of the products. For our purposes the latter comes down to the following priority categories: (1) the product does not need a visit now, (2) a visit to the product may very well prevent an emergency call from the product or customer, (3) an emergency call from the product or customer coming down to the fact that a visit to the product is necessary immediately. Deriving these three categories of priorities from product status information is determined by e.g. technical aspects and service level agreements. Obviously, if preventive service visits are ignored (category 2), routing is steered by emergency calls. It is the aim of this example to assess the value of using the monitoring information and preventing emergency calls in order to improve the efficiency of routing. In order to assess the value of status information, let us formalize the business case to some extent.

Each customer is characterized by its geographic position x and its product status s. We assume that each service vehicle will service Q customers per day. The costs that we focus on in this example simply are the travel costs linear with the length of the route along the Q customers. The aim of using installed base information here is to reduce the average length of the service route.

A tour length L is defined by the order in which customer locations $x_1,\ldots,x_Q$ are serviced, namely $L = d(x_0, x_1) + d(x_1, x_2) + \ldots + d(x_Q, x_0)$, where x_0 is the location of the service center from which the service vehicle departs and where it arrives at the end of its service tour, and where d(x,y) is the travel distance between locations x and y. An optimal tour length is obtained by servicing the customers in an optimal order. This optimal tour will be denoted by $L^*(x_1,\ldots,x_Q)$. Finding such an optimal tour is the aim of the Traveling Salesman Problem. It is also possible to settle for a near-optimal tour using a heuristic. For details, see (Bramel and Simchy-Levi, 1997).

Information on the installed base should enable an additional optimisation step. Each day, it should provide a customer pool of N customers, where N > Q, from which Q customers are to be selected. This large pool of customers use products that either need immediate servicing (category 3) or that may need servicing in the near future (category 2). Given N customers $x_1,\ldots,x_N$, one needs to find Q customers $x_{k(1)},\ldots,x_{k(Q)}$ from this population that optimises $L^*(x_{k(1)},\ldots,x_{k(Q)})$. In the selection of the subsample of Q customers, there may be the restriction that a limited number of customers must be selected anyway, namely category 3, products that require immediate servicing.

The foregoing optimisation procedure may seem an almost impossible task. First of all, finding an optimal tour is not easy (the problem is NP-hard). Secondly, all subsamples of Q customers from N customers need to be checked. However, using heuristics to obtain near-optimal routes may resolve computational problems. Further, some first results indicate that efficiency improvements are considerable, as the following numerical experiment shows. This may result in sufficient benefits without checking all possible subsamples.

A simple example in which Q = 2, so that $L^*(x_1,x_2) = d(x_0, x_1) + d(x_1, x_2) + d(x_2, x_0)$, and an efficiency parameter that is the quotient of the sum of the customer distances $d(x_0, x_1) + d(x_0, x_2)$ and the tour length, provides the following. It is not difficult to see that the worst case scenario results in an efficiency parameter value of ½, and that the best case scenario results in an efficiency parameter value of 1. The following diagram (Figure 5) shows simulation results of N customer draws from a homogeneous distribution of customers from which the optimal tour involving Q = 2 customers is selected. As N increases, the average efficiency converges to 1. This can be proved rigorously, also for more general instances.

This example focuses on transport costs savings in operations. To the authors of this paper it served as an eye opener in the sense that information on the Installed base can be crucial. As can be seen in e.g. (Klausner et al., 1998), a number of issues come up in the use of product status information for decision making. The status of the product needs to be measured and logged in such a way that priority categorization is robust and not error prone.

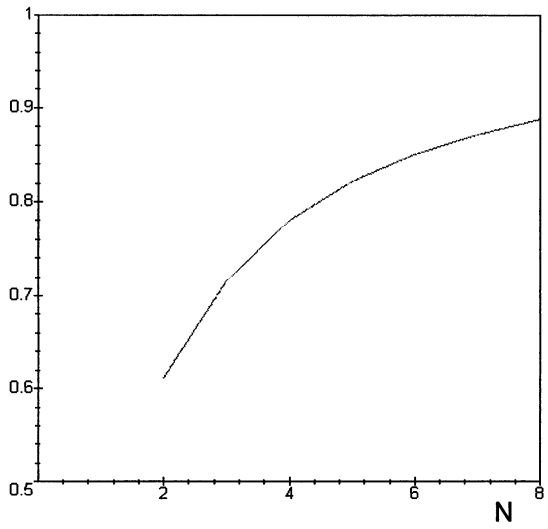


Fig. 5. Average efficiency parameter for different values of N and Q = 2.

The economic value of using installed base information may be delicate and will depend on the value of the product and services offered to the customer and to the other efficiency gains mentioned, and the cost and investments required to monitor the installed base. However, other factors may be of even greater importance. For instance: pro-active servicing also prevents damage and downtime. In situations where not only the defect component incurs cost, but particularly the domino effect through the entire system, the instalment of (expensive) monitoring systems is justified by reducing indirect cost and improving customer service. Moreover, once discarded, valuable information can be obtained on e.g. the condition of the return product/component, generally acknowledged to be a critical factor in value recovery. Using this information, e.g. the internet can consolidate supply and demand of reusables.

Another question is the type of parameters to be monitored. In the reverse channel, this might involve tracking and tracing of the returns and e.g. recovery capacity planning. The installed base is crucial in responsive CSCs since it serves both as a sink and a source. On the return side the aim is to collect reliable information on volume, quality, timing, location and composition of returns. This means in concrete terms:

- Supplying reliable product information (disassembly BOMs, hazardous materials, yield factors)

- Tracking and tracing locations and numbers of products and systems
- Mean time between failures (MTBF): hence return rate
- Phase outs of systems or components
- Supporting product acquisition (consolidation and pricing mechanisms)
- Preventing failures and downtime (diagnostics)

One the (re) use side it is important to determine secondary demand hence volume, composition, quality, location and timing required. In other words:

- Monitoring actual use and customer profile
- Reconfiguration and upgrade planning
- MTBF, hence demand rate
- Recovery BOMs (replacement factors, commonalities)
- Support secondary sales (trading, consolidation)
- Mass balances and environmental impact reporting
- Re-engineering through feed back of failure information to designers

6 Concluding Remarks

This paper aims to contribute to the CSC fields body of knowledge by taking a value creation perspective.

Value contained by returns involves negative externalities, intrinsic value or time based value. Depending on this, a control, efficient or responsive CSC is needed. We discussed a number of cases, and drew both practical and modelling conclusions. Finally we gave a theoretical example of the potential of these concepts in a routing problem that is well studied in operations research.

Future research will focus on a more detailed specification of the installed base control variables, preferably differentiated per type of CSC, and the development of formal models to assess and optimize value recovery through application of IT.

References

Aken, J. E. van (1994): De bedrijfskunde als ontwerpwetenschap, de regulatieve en de reflectieve cyclus, *Bedrijfskunde*, 66-1, pp.16-26 (in Dutch)

Bloemhof, J. / Fleischmann, M. / van Nunen, J. (1999): Reviewing Distribution Issues in Reverse Logistics. in: Grazia Speranza, M. / Stähly, P. (eds.): New Trends in Distribution Logistics, Lecture Notes in Economics and Mathematical Systems 480, Springer Verlag, pp. 23-44

Bramel, J. / Simchy-Levi, D (1997): The Logic of Logistics, Springer Verlag, New York

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

Fisher, M. L. (1997): What is the right supply chain for your product? – a simple framework can help you figure out the answer, *Harvard Business Review*, march-april issue, pp. 105-116

Fleischmann, M. / Krikke, H. / Dekker, R. / Flapper, S. D. (2000): A Characterisation of Logistics Networks for Product Recovery, *Omega, the international journal of Management Science*, 28-6 , pp. 653-666

Giuntini, R. / Andel, T. (1995): Reverse Logistics role models, *Transportation and Distribution*, April, pp.97-98

Guide, D. / Van Wassenhove, L. (2002): Closed loop supply chains. in: Klose, A. / Speranza, M. G. / Van Wassenhove, L. (eds.): Quantitative Approaches to Distribution Logistics. Lecture Notes in Economics and Mathematical Systems 519, Springer Verlag, Berlin, pp.47-60

Guide, D. / Van Wassenhove, L. (2000): Managing product returns for remanufacturing, *Production and Operations Management*, 10:2

Klausner, M. / Grimm, W. / Henderson, C. (1998): Reuse of Electric Motors in Consumer Products, Design and Analysis of an Electronic Data Log, *Journal of Industrial Ecology*, issue 2-2, pp. 89 –102

Kokkinaki, A. / Dekker, R. / de Koster, R. / Pappis, C. / Verbeke, W. (2001): From E-trash to E-treasure: How value can be created by new e-business models in reverse logistics, *Proceedings of the Euro conference*, Rotterdam

Krikke, H. / LeBlanc, I. / van de Velde, S. (2004): Product modularity and the design of closed loop supply chains, to appear in *California ManagementReview,* winter issue 2004

Krikke, H. / Kokkinaki, A. / van Nunen, J. (2003): IT in closed loop supply chains, in:, Guide, D. / Van Wassenhove, L. / Dekker, R. / Corbett, C. (eds.): Business aspects of Closed Loop Supply Chains. Carnegie Bosch Institute, Pittsburgh, pp.255-290

Kuik, R. / van Nunen, J. / Gerrits, J. / Hogenboom, M. (2004): OMRON: Fast Repair and Return" in:, Flapper, S. D. J. / van Nunen, J. A. E. E. / Van Wassenhove, L. N. (eds): Managing Closed-Loop Supply Chains. Springer Verlag, Berlin, to appear

Marien, E. J. (1998): Reverse Logistics as a competitive tool, *Supply Chain Management Review*, Spring 1998, pp.43-52

Rogers, D. S. / Tibben-Lembke, R. S. (1998): Going backwards: Reverse Logistics trends and practices, Executive council of Reverse Logistics, www.rlec.org, 1998

Spengler, T. (2002): Management of Material Flows in Closed-loop Supply Chains, Decision Support for Electronic Scrap Recycling Companies, presented at the WEEE workshop at INSEAD, Oct.17-18, 2002

Stock, J. (2001): "Avoiding the seven deadly sins of Reverse Logistics", *Material Handling Management*, 56:3, pp.5-11

Van Galen Last, J. (1998): Product Return Flows & Physical Distribution Complaints. Unpublished Master Thesis, Rotterdam School of Management

Vorst, J. van der (2000): Effective Food Supply Chains, Ph.D thesis, Wageningen University

Workshops at INSEAD, Closed Loop Supply Chains II, October 15, 16, 17 and Waste on Electrical and Electronical Equiment, October 17, 18, 2002, INSEAD, Fontainebleau, France

Reverse Logistics: A Review of Case Studies

Marisa P. de Brito[1], Rommert Dekker[1] and Simme Douwe P. Flapper[2]

[1]Econometric Institute, Erasmus University Rotterdam, The Netherlands
[2]Faculty of Technology Management, Technische Universiteit Eindhoven, The Netherlands

Abstract. Reverse logistics deals with the processes associated with the flows of products, components and materials from users/owners to re-users. This paper provides a review and content analysis of more than sixty case studies on reverse logistics. The case studies deal with issues such as the structure of the networks, the relationships between the different parties involved on the networks, inventory management, planning and control, and information technology. The analysis concerns the following questions: what is returned, why do these return flows exists, i.e. what are the drivers/reasons for these flows, how are these return flows being recovered, and finally, who is involved? We end with a summary and suggestions for further research.

Key words: Reverse logistics; Case studies; Content analysis.

1 Introduction

Reverse logistics, dealing with the physical flows of products, components and materials flowing from users/owners to re-users, is a growing field both in practice and in the academic world. In order to get insight in how reverse logistics are dealt with in practice, field studies and surveys are useful. Examples of the latter are presented in (Rogers and Tibben-Lembke, 1999; Guide, 2000).

This paper gives some idea about the diversity of logistics practices via a content analysis for more than 60 case studies. This analysis is useful for both practitioners and academics, by giving insight in how companies are dealing with reverse logistics, which are the trade-offs that play a role during decision making, how these decisions are supported and so on. Quite some case studies on reverse logistics have been described in the literature, dealing with different industries, recovery options and drivers, see e.g. (Kopicki et al., 1993; Kostecki, 1998; Lund, 2001; Flapper et al., 2004). However, there are many more case studies available in literature that are scattered over journals for very different research communities. Besides, in a number of countries in the forefront of reverse logistics, a substantial number of case studies has been published in languages other than the lingua franca and is therefore not accessible for the majority of the research community. The main purpose of this paper is to give an overview of the above case studies.

The remainder of the paper is organized as follows. In Section 2 we firstly define what we mean by case study and describe the methodology used for finding and classifying the case studies presented in this paper. Next, in Section 3, we

B. Fleischmann et al. (eds.), *Distribution Logistics*

provide overall statistics for the set of cases discussed. After that, in Sections 4 to 8, we discuss the case studies that we found in detail, and present observations, propositions and research opportunities. Finally, in Section 9, we end with a summary of our main findings and our main conclusions.

2 Methodology

By case study we mean an in-depth analysis of a practice, leading to an understanding of a situation within its context (Stake, 1995). This definition rules out mere examples short on details, as well as the sole description of business practice often found in professional journals. The latter does not mean that we think that professional journals are not useful for learning about reverse logistics. However, publications in these journals are more likely to be biased than publications in scientific literature, for instance because often they are also used as a marketing tool. We used the following on-line sources: Science Citation Index Expanded (http://www.sciencedirect.com/) of the Institute of Scientific Information, and ABI/INFORM Global (http://www.proquest.umi.com) of the Bell & Howell Information and Learning Company, using the key words listed in Table 1 and (a combination of) the words: logistics, planning, control, transport, inventory, capacity, production, information.

Table 1. Key-words used for the literature search

Asset recovery	Post-consumer	Repair
By-products	Producer responsibility	Repairable
byproducts	Product ownership	Resale / re-sale
Containers	Product recovery	Resell / re-sell
Co-products	Product stewardship	Return (includes commericial returns)
coproducts	Reassembly	Reuse/ re-use
Core	Rebuild	Reutilisation
Defects	Recalls	Reusable
Defective	Reclaim	Reverse logistics
Disassembly	Reclamation	Rework
Dismantling	Reconditioning	Salvage
Disposal	Re-consumption	Secondary (market, materials)
Downgrading	Recovery (product, resource, asset)	Separation
Energy recovery	Recycling	Source reduction
Environment	Refill	Take back
Garbage	Refillable	Upgrading
Gate keeping	Refurbishing	Value recovery
Green logistics	Remanufacturing	Warranty
Material recovery	Repack	Waste
Obsolete (stock)		
Outlet		
Overstock		

Moreover, we went through some other sources dedicated to reverse logistics: the (Dutch) Handbook of reverse logistics, and the proceedings related to the

Reman conferences organized by APICS (The American Production and Inventory Control Society), see (http://www.apics.org/), and the proceedings related to the conferences on Electronics and the Environment, organized by IEEE (Institute of Electrical and Electronics Engineers), see (http://www.ieee.org/). The main search took place in early 2001 but continued till the end of 2002. The case studies are structured according to a framework for reverse logistics presented in (De Brito and Dekker, 2004). Accordingly, for each case study we collected data on

- What: the products entering the reverse logistics network (product-in's) and the products leaving the reverse logistics network (product-out's)
- How: the main recovery process used
- Who: the actors and their function in the reverse logistics network (supplier (i.e. the owners/users of the product-in's), collector, processor, customer and initiator);
- Why: the driving forces for the suppliers of the product-in's and the initiator of the reverse logistics activities.

The above information is summarised in a number of tables, where each table gives the following general case information: the reference and the geographical area of the case study, the product-in, the main recovery process used, the product-out, the supplier, the collector, the processor, the customer (re-user), the initiator, the reason(s) for the sender to take part, the driver(s) for the initiator, where an empty cell in a table indicates that the corresponding information is not given in the reference.

Hereafter a summary is given of the different sub classifications used for describing the different cases, based on (De Brito and Dekker, 2004).

Products:

- civil objects (like buildings, dikes, bridges, roads)
- consumer goods (like furniture, TV sets, cars)
- industrial goods (like trucks, machines)
- ores, oils and chemicals
- other materials (like glass, paper pulp)
- distribution items (like bottles, crates, pellets)
- spare parts

Recovery processes:

- direct recovery, i.e. recovery without any major processing
  - re-sale
  - re-use
  - re-distribution
- recovery requiring processing
  - repair (i.e. making products working again, bringing them to working condition)
  - refurbishing (i.e., product upgrading)

- remanufacturing (i.e., the recovery of products to an as new level)
- (parts) retrieval (i.e., the recovery of a selected number of parts from products)
- recycling (i.e., the recovery of materials from products)
- incineration
- (proper) disposal.

Actors:

- forward supply network actors (such as manufacturers, wholesalers, retailers, service providers)
- specialized reverse logistics actors (such as recyclers, independent remanufacturers)
- governmental entities (such as European Union and national governments)
- opportunistic players (such as charity organisations)

Return reasons:

- Manufacturing returns, i.e., returns related to the execution of production processes, including
  - raw material surplus
  - quality-control returns
  - production leftovers
  - by-products
- Distribution returns, i.e., returns related to the distribution of production to (potential) customers, including
  - product recalls,
  - B2B commercial returns,
  - stock adjustments,
  - distribution items
- Market returns, i.e., returns from the users of products, including
  - B2C commercial/reimbursement returns
  - warranties
  - service returns (repairs and spare parts)
  - end-of-use returns
  - end-of-life returns

Drivers:

- Economics (direct and indirect profits related to reduced production costs, green image, market protection, improved customer/supplier relations, etc)
- Legislation
- Corporate Citizenship

The main idea behind de Brito and Dekker's framework is that a specification of the why, what, how and who using the above classification defines and explains the various ways in which reverse logistic activities are done. Moreover, the hypothesis is that a specification of the why and the what determine the processes (the how) and the actors (the who's).

In the context of reverse logistics, companies have to make several strategic, tactical and operational decisions. At strategic level, the collection network has to be designed. At tactical level, the relationships with partners have to be developed. At operational level, inventories have to be managed and activities have to be planned and controlled, see also (Ganeshan et al., 1999; Fleischmann et al., 1997). Based on the above, we discuss the case studies according to the following decision-making focus: Network Structure, Relationships, Inventory Management, and Planning and Control. As usual, information and communication technology plays an important role. For this reason, we also give an overview of Information and Technology (IT) for reverse logistics. If a case study focuses on more than one decision area, we discuss it in more than one section.

3 Statistics

Using the United Nations' classifications for industry (see http://esa.un.org), it turns out that roughly 60% of the cases concern manufacturing, 20% concern wholesale and retail trade, and about 10% concern construction. We have also found cases related to the classes transport and communication, public administration and defense, and other community, social and personal service activities.

Using the United Nations' classification for products (see http://esa.un.org), we observe that almost half of the cases deal with metal products, machinery and equipment. About 30% of the products being processed are other transportable goods like wood, paper and plastic products, whereas about 20% concern food, beverages, tobaccos, textiles and apparel. Less than 10% of the cases are related to the class Ores and minerals.

The majority of the cases are from Europe. In fact, we report on 1 case from Central- America, 1 case from South-America, 2 cases from Asia, 17 cases from North-America and more than 40 cases from Europe, where some cases are related to more than one geographical area. The unequal distribution of cases over geographic areas corresponds to the unequal past development of reverse logistics research in the different continents.

4 Case Studies on Reverse Logistics Network Structures

Main activities in reverse logistics are the collection of the products to be recovered, their processing and the redistribution of the processed goods. Although at first sight, these activities seem to resemble the activities in standard forward production-distribution networks, there are also some differences.

Reverse logistics networks often have many suppliers (owners/users of products) from whom products are collected, far more than in many forward production-distribution networks. Moreover, the condition and configuration of the returned products is often uncertain, the packaging of the returned products is generally more problematic, the cooperation of the suppliers is much more needed and often the collected products have a relatively low value. In case a new network for reverse logistics has to be setup, decisions have to be made with respect to the number of echelons in the network, the number and locations of intermediate depots, the use of drop points for collection, the integration with the forward network of new products, and the financing of the network. See also (Fleischmann et al., 2004).

4.1 The case studies

We found 26 case studies on this subject, presented in Table 2. Hereafter, we discuss the cases in more detail, using the classification based on the recovery process used, which was also done in (Fleischmann et al., 2000). The most often described recovery process is direct recovery, either re-distribution or re-sale (12 cases), followed by recycling (11) and remanufacturing (3 cases).

Networks for re-use/re-distribution/re-sale

The cases that we found concern on the one hand commercial returns to retailers and on the other hand distribution items such as bottles, crates, pallets and containers.

De Koster et al. (2001) compare the return handling operations of three food retailers, three department chains and three mail-order companies. In total these are nine cases.

Kroon and Vrijens (1995) discuss the design of a network for reusable containers. The issues addressed by them are the role of the different actors, the economics of the system, the costs allocation to the different actors, the amount of containers needed, and the locations of the depots for the containers. Del Castillo and Cochran (1996) study the integral planning of production, product distribution and collection of re-usable containers used for distributing the products. They apply their model to the re-usable bottles used by a soft drink company in Mexico. Duhaime et al. (2000) discuss the distribution and collection of returnable containers by Canada Post.

Networks for remanufacturing

Typically, remanufacturing concerns complex equipment with many parts. Usually, remanufacturing is labour-intensive, requiring much testing. Fleischmann et al. (2000) distinguish further between networks setup by OEMs (Original Equipment Manufacturers) and networks setup by independent third parties, because in the latter case integration with the original production-distribution network is hardly possible.

Krikke et al. (1999a) discuss the remanufacturing of copiers. The authors examine the economic consequences of two alternative locations for the remanufacturing facility: one coinciding with the location of the manufacturing facility and one in a cheap labour country. Meijer (1998) discusses Canon's remanufacturing network for printers, scanners, copiers and faxes in the Netherlands, Belgium and Luxemburg. Dijkhuizen (1997) discusses the remanufacturing network of IBM. He tackles the question: where to remanufacture the products of IBM in Europe: in each country or at one central location?

Networks for recycling

Two types of networks can be distinguished in this context: private and public networks. The same classification has been used by Fleischmann et al. (1997) and Goggin and Browne (2000). We identified eight cases on public networks and three on private.

Private networks. Louwers et al. (1999) discuss the setup of a carpet recycling network for part of Europe. Realff et al. (2000) discuss a similar network in the USA. Spengler et al. (1997) discuss two networks: one for the recycling of building waste, and one for recycling the by-products in the German steel industry. The authors pay among others attention to the cooperation between companies.

Public networks. There are several papers describing the setup and organization of public recycling networks in the Netherlands. Bartels (1998) describes the Dutch nation wide network for battery recycling. In the Netherlands since 1995, all manufacturers and importers of portable batteries with a weight of less than one kg have to collect and recycle as much as possible the batteries that they sell in the Dutch market. For this purpose, the above mentioned companies founded the Stibat organisation to take care of this on behalf of them. De Koster et al. (2000) describe the Dutch nationwide network for the collection and processing of white goods, paying special attention to big white goods like refrigerators. Buyers of new white goods have to pay a disposal fee, which is used for financing the network. Van Notten (2000) describes the Dutch glass recycling network. In order to cope with the Dutch legislation, the glass industry in the Netherlands founded the Glass Recycling Foundation. The authors discuss the collect and pickup systems for the collection of glass from households.

Van Burik (1998) describes the Dutch nationwide network for dealing with car wrecks. Car manufacturers, importers, car recyclers and other parties together established a special organisation called Auto Recycling Netherlands (ARN) to fulfil legal requirements. Like in the case of white goods discussed earlier, in the Netherlands buyers of a new car have to pay a fee to finance the system. Barros et al. (1998) discuss the network for the recycling of sand from building waste in the Netherlands. Because the Dutch government aims to increase the percentage of sand that is recycled, from 70% at the moment to 90% in the near future, a group of construction waste processors wants to improve the sand recycling network, by the number and locations of depots where the sand is stored waiting to be recycled and the depots where the actual recycling will take place.

Table 2. Case studies focusing on Network Structures

Reference / Geographical area	Product-in	Process	Product-out	Supplier (owner / user)	Collector	Processor	Customer	Initiator	Reason return Sender (owner / user)	Drivers initiator
Barros et al. (1998) / Europe	building waste	recycling	sand	waste processors	consortium building waste processors	consortium building waste processors	building industry	consortium building waste processors	end-of-life (waste disposal)	legislation
Bartels (1998) / Europe	batteries	recycling	materials	households, companies	municipalities, retailer chains, schools	specialized companies		dutch master organization of importers of batteries	end-of-use (environmental responsibility)	legislation
Chang et al. (2000) / Asia	house-hold waste	recycling		households	public authority environmental protection bureau	public authority environmental protection bureau		Kaohsiung's city government	end-of-life (waste disposal)	(public) responsebility
Del Castillo and Cochran (1996) / Central-America	bottles	redistribution	bottles	consumer soft drinks	retailers		soft drinks producer	soft drinks producer	end-of-use (plus deposit fee)	economic (cost savings)
De Koster et al. (2000) / Europe	large white goods	recycling	materials	households	municipalities, retailers, 3rd party logistics service providers	recycling companies (CoolRec, HKS)	materials processing companies	suppliers of household appliances, producers, importers of browngoods	end-of-life (no disposal costs)	legislation
De Koster et al. (2002) / Europe (3 cases)	products, packaging, distribution items	redistribution	products, packaging, distribution items, waste	customers, super-market stores	supermarket stores and distribution centers	super-market chain	supermark-et chain, collectors, processors, suppliers	supermarket chain	functional, commercial & end-of-life (less disposal costs)	economic (less disposal costs)

Table 2 continued

De Koster et al. (2002) / Europe	shoes, sports attributes, waste, advertise-ment materials	redistribu-tion	products, packaging, distribution items, waste	customers, super-market stores	supermarket stores and distribution centers	super-market chain	supermarket chain, special outlets	supermarket chain	reimburse-ment	economic
De Koster et al. (2002) / Europe	products, coat hangers, racks	redistribu-tion	products, coat hangers, racks, waste	customers, supermarket stores	supermarket stores and distribution centers	supermarket chain	supermarket chain, special outlets	supermarket chain	reimburse-ment	economic
De Koster et al. (2002) / Europe	products, advertise-ment materials, containers	redistribu-tion	products, containers, waste	customers, supermarket stores	supermarket stores and distribution centers	supermarket chain	supermarket chain, special outlets	supermarket chain	reimburse-ment	economic
De Koster et al. (2002) / Europe	consumer goods	redistribu-tion	consumer goods	customers	mail order company	mail order company	same market	mail order company	reimburse-ment	economic, legislation
De Koster et al. (2002) / Europe	consumer goods	redistribu-tion	consumer goods	customers	mail order company	mail order company	same market	mail order company	reimburse-ment	economic, legislation
De Koster et al. (2002) / Europe	cloths, small appliances	redistribu-tion	cloths, small appliances	customers	mail order company	mail order company	same market	mail order company	reimburse-ment	economic, legislation
Dijkhuizen (1997) / Europe	defective parts	remanu-facturing	remanu-factured parts	customers IBM	IBM	IBM	customers supplying defective parts	IBM	service returns	economic (market protection quality feedback) legislation
Duhaime et al. (2001) / North-America	distribution items	redistribu-tion	distribution items	Canada Post	Canada Post	Canada Post	Canada Post	Canada Post	functional returns	economic

Table 2 continued

Kleineidam et al. (2000) / Europe	paper	recycling, incinera-tion	materials, energy	households, companies	mainly non-profit organisations	recycler, special collector	pulp industry		end-of-life (reduction waste disposal costs)	economic
Kroon and Vrijens (1995) / Europe	distribution items	redistribu-tion	distribution items	customers	Nedlloyd	Nedlloyd	customers		end-of-use (incentive: deposit fee)	economic
Krikke (1999a) / Europe	copiers	remanu-facturing	remanu-factured copiers	Oce national sales organisations	Oce	Oce	Oce national sales organizations	Oce	end-of-use (incentive: fee)	economic
Louwers et al. (1999) / Europe	carpets	recycling	fibers, filling materials for roads, dams etc	households, companies (e.g. involved in floor covering)	companies involved in floor covering, municipalities, special organization	organization for sorting, fiber producers, cement industry	customers for fibers, cement industry, road builders	carpet industry, including their fiber suppliers	end-of-use/ end-of-life	economic (image); legislation (expected)
Meijer (1998)/ Europe	scanners, printers, copiers, toner cartridges packaging	remanu-facturing, recycling	remanu-factured machines, materials	households, companies	dealers, Third party logistics service providers	Canon, Fr (toner cartridges), Canon Scotl (copiers)	Canon	Canon	end-of-use/ end-of-life (incentive: no disposal costs)	economic (green image)
Realff et al. (2000) / North-America	Carpeting Mat.	recycling	nylon fibres	business customers	carpet dealers	DuPont		DuPont	end-of-life (waste disposal)	economic (market value)
Spengler et al. (1997) / Europe	by-products steel pro-duction	recycling	materials	steel industry	steel industry	steel industry	steel and other industry	industry, government	manufactur-ing (leftovers); disposal cost savings	
Spengler et al. (1997) / Europe	demolition	recycling	materials	demolition companies	recycler	recycler		government	end-of-life (disposal cost savings)	

Table 2 continued

Van Burik (1998) / Europe	car wrecks	recycling	materials, waste	car owner	certified disassembly companies	selected recycle companies		car producers, importers in Netherlands	end-of-life (incentive: no disposal costs)	legislation, economic
Van Notten (2000) / Europe	glass	recycling	materials (input for glass industry)	households, companies	specialized companies	glass recycling companies	glass industry	glass collectors, recyclers, producers	end-of-life (incentive: no disposal costs)	legislation

Kleineidam et al. (2000) consider the structure of the recycling network of the paper industry in the Netherlands. Companies selling paper and cardboard to the Dutch market, setup the Dutch Paper Recycling Cooperation. This cooperation takes care of the collection of paper, its processing into pulp that is the raw material for the paper industry. The authors focus on the network for the collection of paper from households run by non-profit organizations. The authors study the dynamic behaviour of the network, analysing the consequences of incineration costs and paper taxes. Chang and Wei (2000) discuss the location of the recycling drop-off stations in the recycling network for household waste in the city of Kaohsiung in Taiwan for which the Environmental Protection Bureau of the city is responsible.

4.2 Remarks

The reader should have noticed that this section focuses on the how and who. Typologies of the how and who also have been (more or less consciously) employed in previous literature to distinguish reverse logistic networks (see Fleischmann et al. 1999; Goggin and Brown 2000; Fleischmann et al. 2000; Fleischmann et al. 2004). Based on the case studies described above, one can make several conjectures about relations between elements of the framework. For example, all public recycling networks are oriented at avoiding waste. They consider end-of-life returns and are based on cooperation between companies. Cooperation or not between companies is a big issue for the other networks. In some cases it is done, but in other cases not.

So far, the models used for deciding on the locations of reverse logistics processing facilities are based on deterministic integer programming models (Fleischmann et al., 2003). However, often the uncertainty in reverse logistics networks is much higher than in the production-distribution networks for many products. Stochastic programming models may be useful in this context (Listes and Dekker, 2001).

5 Case Studies on Reverse Logistics Relationships

To stimulate/enforce a certain behavior of their partners, parties in the reverse chain may use various incentives. In this section we will discuss our findings with respect to the incentives that are used in practice to stimulate/enforce a desired behavior of partners in the context of product recovery. Thereby a distinction should be made between 1) incentives that may be used to influence the supply of goods to a company in the context of product recovery, and 2) incentives that may be used to influence others to accept the goods a company wants to get rid of. To be concrete: a producer of toner cartridges may be interested in incentives for getting back its cartridges, whereas a company buying chemicals in kegs for producing its products, may want these kegs to be collected and environmentally consciously processed by its supplier in order to avoid the high costs for disposing the kegs once they are empty.

Defining incentives to stimulate/enforce the desired behaviour of partners both outside and inside a company requires insight into the alternatives for these partners with respect to the products to be recovered and the costs (time, money, space) and benefits related to each of these alternatives. Clearly the incentive should be relevant, where creating a win-win situation might not be enough.

Incentives are used to influence 1) the quantity, 2) the quality (configuration and condition), and 3) the moment of supply. These three factors determine to a large extent the possibilities for recovery and reuse.

5.1 The case studies

Table 3 gives the general data for the case studies that we found where special attention is paid to the above mentioned incentives. We found eight different types of incentives in literature.

Deposit fee

This fee may concern the product itself or the item used for its distribution like a bottle, box or pallet. An example of the former is the deposit fee that has to be paid when renting a car, whereas an example of the latter are the deposit fee on beer bottles from glass and PET bottles used for the distribution of soft drinks, see e.g. (Vroom et al., 2001).

Take back

There are different types of take back options offered in practice. A company may offer to collect its products from its customers when these customers want to dispose them, either for free, or for a lower price than the customers would have to pay else. An example of the latter is the take back program of Rockwool Lapinus B.V., subsidiary of the Danish Rockwool company, in the Netherlands (Wijshof, 1997). Sometimes, at the moment that a product is sold, the buyer is offered the possibility to sell the product to the seller for a preset price during some time after the product has been bought if the returned product fulfils some preset requirements at the moment of return based on the use of the product, like kilometers driven, and expected possibilities for selling the returned product. This holds for instance, for the Ford Options program (http://www.ford.nl). Numerous examples of buy back options for unused products are presented, see e.g. (Tsay, 2001), where no explicit attention is paid to what happens to these products hereafter.

Fee

A fee is paid when a product is delivered for recovery. Usually the fee depends on the condition and configuration of the product delivered, but sometimes also on the moment that a product is delivered because this may determine the possibilities to reuse (parts of) it. Well-known examples of companies using fees

Table 3. Case studies focussing on Relationships

Reference / Geographical area	Product-in	Process	Product-out	Supplier (owner / user)	Collector	Processor	Customer	Initiator	Reason return Sender (owner / user)	Driver initiator
Bartel (1995) / North-America	toner cartridge	remanu-facturing, (parts) retrieval	toner cartridges or parts	customers of Unisys	US postal services		customers Unisys	Unisys	end-of-use (fee as an extra incentive)	economic (to offer recovered toner cartridges)
Driesch et al. (1998) / Europe	car engines	remanu-facturing, refurbishing	car engines	owners of a Mercedes Benz (MB) with an MB engine	Mercdes-Benz dealers	Daimler Chrysler	owners of an MB car	Daimler-Chrysler (Mercedes-Benz)	Service (trade-in)	economic (customer relations)
Faria de Almeida and Robertson (1995) / Europe	batteries	recycling	materials	battery users	shopping malls, stores, mobile collection units, recycling points	specialists	customers materials	City of Leicester (UK)	end-of-use	
Farrow et al. (2000) / North-America	cardboard boxes	redistribu-tion	Cardboard boxes	Walden Paddlers	New England Foam of Windsor		New England Foam of Windsor	Walden Paddlers	end-of-use (reduction waste disposal costs)	corporate citizenship
Guide and Van Wassenhove (2000) / North-America	toner cartridge	remanu-facturing	toner cartridges	customers of Xerox	US Postal Services		customers of Xerox	Xerox	end-of-use returns (fee as an extra incentive)	economic
Guide and Van Wassenhove (2001) / North-America	cellular phones	remanu-facturing	cellular phones	cellular airtime providers	ReCellular, third party	ReCellular		ReCellular	end-of-use (fee as an extra incentive)	economic

Table 3 continued

McGavis(1994) / North-America	toner cartridges	recycling	materials	user cartridge	UPS	specialists for different parts	HP	HP	end-of-use returns (donation by HP to WWF as an incentive)	economic
Vroom et al. (2001) / Europe	PC bottles, crates, pallets	redistribution	PC bottles, crates, pallets	households, supermarkets	supermarkets	Campina	Campina	Campina	end-of-use returns (deposit fees as an incentive)	economic
Wijshof (1997) / Europe	Rockwool produced by Rockwool Lapinus	recycling	materials (secondary input production rock wool)	building companies, demolishers, other buyers	BFI, Sovabo	Rockwool Lapinus	customers Rockwool Lapinus	Rockwool Lapinus	end-of-life (less disposal costs than other disposal options)	economic (green image)
Yender (1998) / Europe	batteries	recycling, reloading	materials, batteries	users batteries	UPS, retailers, municipalities	different specialists		rechargeable battery recycling corporation	end-of-life / use	

to stimulate the supply of products for recovery are car brokers and "second hand" shops. Other examples are Varta, the German battery manufacturer, in the UK paying 50 pence for every returned rechargeable battery sent to a collection point (Faria de Almeida and Robertson, 1995), UNISYS, paying for each toner cartridge returned (Bartel, 1995), and ReCellular, buying used cellular phones (Guide and Van Wassenhove, 2001). Sometimes, not the supplier of a product receives a fee for each product returned, but a non-profit organisation that may or may not be chosen by the supplier. This incentive was used by Hewlett Packard in order to get her toner cartridges back after an attempt to refer to the environmental consciousness of customers did not result in a satisfying number of returns (McGavis, 1994), and by Tesco to get cellular phones (http://www.tesco.com).

Trade-in

One can only get a new copy of a product if another copy is returned. This incentive is among others used by Daimler-Chrysler for the engines that they produce for Mercedes-Benz passenger cars and small vans (Driesch et al., 1998). Owners of a Mercedes-Benz (MB) passenger car or small van with an MB engine can go to an authorised MB dealer in order to have their present engine be replaced by a reconditioned engine. The MB dealer removes the engine and sends it to the central parts DC of MB. From this DC, the reconditioned engine is sent to the dealer where the engine is available within 24 hours. The MB-dealer puts the reconditioned engine in the MB passenger car or van, after which the car can be used again by its owner.

Easy and simple method of supply

Two main supply systems can be found in practice: 1) pickup systems where (parts of) products to be recovered are collected at the locations where they are disposed, and 2) bring systems where the disposer has to bring the products to a certain location (Kopicki et al., 1993). In practice, usually combined systems are found, like in the case of glass containers where the households have to bring the glass to a container that is emptied by a collector who brings the glass to a processor (Lund, 2001). Tucker et al. (2000) study the optimal picking frequencies to achieve a certain collected quantity. Some suppliers for toner cartridges, including UNISYS, deliver their cartridge in a box that can be returned for free either by mail (mixed bring-pickup system) (Bartel, 1995) or via a third party logistics service provider (pickup system) as done by Hewlett Packard (McGavis, 1994). The same incentive is used by some manufacturers of batteries (Yender, 1998).

Timely and clear information about the reverse logistics activities

How important this incentive is, is illustrated by a pilot system for the collection of different types of batteries in Denmark and Germany, where it turned out that it was for a number of batteries too difficult to make suppliers clear which type of battery they have (Faria de Almeida and Robertson, 1995).

Environmental responsibility

The idea is to appeal to the environmental consciousness of people. This incentive usually requires a lot of advertising, and is in general not very reliable as is illustrated by the collection of toner cartridges by Hewlett Packard (McGavis, 1994).

Power

As always, power can be used to force desired behaviour. An example is Walden Paddlers, using her power as a customer to force New England Foam of Windsor to take back the cardboard boxes that this supplier uses to distribute foam foot braces and seat pads to Walden Paddlers (Farrow et al., 2000).

5.2 Remarks

From the above, it is clear that a variety of incentives are used in practice to influence the behavior of partners in reverse logistics networks. Some incentives make up part of sales contracts, like the buy back option offered by Ford (http://www.ford.nl) and deposit fees. Other incentives, like trade-ins, require the customer to buy another product, as applies to the engine trade-in offered by Daimler-Chrysler (Driesch et al., 1998). There are also incentives not directly coupled to a selling activity, like a gift to a non-profit organisation as used by Hewlett Packard (McGavis, 1994) and Tesco (http://www.tesco.com). It seems that only deposit fees are specific for product recovery. The other incentives are also used to attract customers in general.

Although all the above mentioned incentives can be found in practice, as far as we know, no models exist to support the decision which incentive to use in which situation. With respect to the values of the different parameters related to an incentive, some research is available. Klausner et al. (1998) and Klausner and Hendrickson (2000) present a mathematical model that might be used for estimating the buy back price. Guide et al. (2001) present a mathematical model to determine the optimal acquisition price for products from the field as well as the selling price of these products. However, in the above two models the time aspect is neglected, i.e. a steady state situation is considered. There is quite a lot of literature on sales contracts with return options for unused products, including (Tsay, 2001; Anupindi and Bassok, 1999; Corbett and Tang, 1999; Lariviere, 1999; Tsay et al., 1999). The prime focus of the latter literature on contracts with a return option is on what may be gained by both sellers and buyers by allowing buyers to order more under certain return options, where a fixed sales price is assumed for the products that are taken back by the seller.

6 Case Studies on Inventory Management

The cases studies focusing on inventory management within reverse logistics are given in Table 4. They can be classified according to the return reasons (see

Section 2). We did not find cases for all reasons. We found cases for B2C commercial returns, service returns, end-of-use returns and end-of-life returns. Omissions can be explained as follows. Manufacturing returns are often treated in production planning contexts. Product recalls are often special events, which are left out of consideration in inventory management. Warranty returns have similar characteristics as repairs. Stock adjustments are somewhat similar with some of B2B commercial returns (bulk returns) and might not even be distinguished from them in the case descriptions. Below we discuss the cases found in detail together with the mathematical models applied.

6.1 The case studies

Commercial returns cases

Commercial returns occur in a B2B or in a B2C setting, where the buyer has a right to return the product, usually within a certain period. The reason behind the return option differs between the cases. In the first setting, the retailer faces the problem of how much he might sell and giving him a buy-back option lowers this risk for him. The returns are likely to be in bulk at the end of the season. In the second case the reason for the return option is that the buyer might not be sure whether the product really meets his/her requirements.

Sanders et al. (2000) describe how the inventories of products are controlled within Wehkamp, a Dutch mail order company, selling all kinds of consumer goods to the Dutch and Belgium market. Two types of products are distinguished: products which are asked for during a very short period of time only (fashion products), which are controlled by using an amended version of the newsboy model taking into account returns; and products that can be sold during a long period of time, which are controlled via a (R, S) policy with variable R and S.

De Brito and Dekker (2003) investigate the distribution of the return lag, i.e. the time between the purchase and the return of an item, and its consequences for inventory management. Three cases are considered, viz. a mail order company, a spare parts warehouse at a petrochemical plant and the warehouse at the center for nuclear research, CERN.

Service return cases

Within service systems (like repair systems) returns occur basically in two ways. First of all, the products themselves may be brought or sent to a center for repair. If the repair is successful, they are brought back; else, a new product or system needs to be bought and the failed one is discarded. Secondly, if one needs a continuous functioning of the product or system, one may directly restore functionality by replacing a part. The failed part is then repaired, after which it will enter the inventory of spare parts. The cases found are described below in detail.

Diaz and Fu (1997) study a 2-echelon repairable item inventory model with limited repair capacity. For several classes of arrival processes they develop analytic expression for the number of items in queue at the different stages of the system. They analyze the impact of the capacity limitation and compare the performance of their approach with an uncapacitated METRIC type of model. Both models are applied to the case of spare parts management at the Caracas subway system. Donker and Van der Ploeg (2001) describe how the optimal stock of reparable service parts of telephone exchanges is determined within Lucent Technologies Netherlands. They use an amended METRIC model, where the service measure is fill rate (i.e. the percentage of demand that can immediately fulfilled from stock) and there is no budget restriction for service parts. Moffat (1992) provides a brief summary of a Markov chain model for analyzing the performance of repair and maintenance policies of aircraft engines at the Royal Air force.

Van der Laan (1997) describes the remanufacturing network of engines and automotive parts for Volkswagen. It is very similar to the engine remanufacturing case related to Mercedes Benz in the previous section. Guide and Srivastava (1998) discuss for an air force depot a method to determine the inventory buffers between the disassembly and the remanufacturing shop, and the inventory buffer between the remanufacturing shop and the reassembly shop.

End- of-use returns cases

This return reason concerns items that are only temporary needed by a user. The product may either be leased, rented or temporary given into the authority of the recipient. The latter is the case with distribution items, that is, products like containers, bottles, railcars and crates, which are used for distribution purposes. The two cases found, viz. Swinkels and van Esch (1998) and Del Castillo and Cochran (1996) primarily concern distribution items. Here the location of the items is a major issue in the inventory decision.

Del Castillo and Cochran (1996) study production and distribution planning for products delivered in reusable containers. Their model includes transportation of empty containers back to the plants. Availability of empty containers is modelled as a resource constraint for the production of the original product. The model is applied to a case study of a soft drink company using returnable bottles.

Swinkels and Van Esch (1998) describe how the optimal stock of refillable beer kegs is determined within Bavaria, a Dutch beer brewery.

Toktay et al. (2000) consider inventory management for Kodak's single use cameras. As the camera acts as a container for the film, one may see this also as a distribution case. Printed circuit boards for the production of these cameras are either bought from overseas suppliers or remanufactured from the cameras returned by the customers via photo laboratories. The issue is to determine a cost-efficient order policy for the external supplies. Major difficulties arise from the fact that return probabilities and market sojourn--time distribution are largely unknown and difficult to observe. The authors propose a closed queuing network model to address these issues. They assess the importance of information on the returns for the control of the network. Rudi et al. (2000) discuss the product recovery actions of the Norwegian national insurance administration. This public entity retrieves no longer needed wheel chairs, hearing aids and similar products

Table 4. Case studies focussing on Inventory Management

Reference / Geographical area	Product-in	Process	Product-out	Supplier (owner / user)	Collector	Processor	Customer	Initiator	Reason return Sender (owner / user)	Driver initiator
De Brito and Dekker (2003) / Europe	laboratory equipment (15 000 sku)	re-use	same product	internal customers	customer brings it back	CERN	internal customers	CERN	commercial returns (reimburse-ment)	economic
De Brito and Dekker (2003) / Europe	fashion, electronics and furniture	redistribution	same product	customers	mail-order-company	mail-order company	customers (same chain)	mail-order-company	commercial returns (reimburse-ment)	economic (marketing) legislation
De Brito and Dekker (2003) / Europe	thousands of spare parts	repair, redistribution	same product	maintenance personnel	maintenance brings it back	refinery (as maintenance)	internal customers		commercial returns (no longer needed)	economic
DelCastillo and Cochran (1996) / Central-America	See Table 2 Network Structures									
Diaz and Fu (1997) / South-America	railway spare parts	repair	repaired spare parts	Caracas Subway	Caracas Subway	Caracas Subway	Caracas Subway	Caracas Subway	repair	economic (cost savings)
Donker and van der Ploeg (2001) / Europe	circuit boards for telephone-exchanges	repair	repaired circuit boards	telephone companies	Lucent Technol-ogies	Lucent Technol-ogies	telephone companies	Lucent Technol-gies	repair	economic (cost savings)
Fleischmann (2000) / Europe	used / unused machines	repair, refurbish-ment	spare parts	business customers / retailers		IBM facilities			commercial (overstocks no longer needed)	economic / legislation

Table 4 continued

Guide and Sri- vastava (1998) North-America	See Table 2 Planning and Control									
Klausner and Hendrickson (2000) / Europe	power tools	remanu-facturing recycling	Remanu-factured power tools, or materials	customers	dealer, logistics provider	specialized facility		manufacturers of power tools	end-of-use	economic (pro-active policy)
Moffat (1992) / Europe	aircraft engine	repair, refurbishing	re-processed aircraft engine	UK Air force	UK Air force	UK Air force	UK Air force	UK Air force	service returns (repairs)	economic (cost savings)
Rudi et al. (2000) / Europe	wheel chairs, hearing aids, etc.	re-use, refurbishing recycling, parts retrieval, landfill	re-processed wheel chairs, hearing aids, etc.	users	Technical Aid Center (TAC)	TAC (in some cases recycling center)	people with handicaps	Norwegian National Insurance Administra-tion	service returns (repairs) end-of-use returns	economic responsibility
Sanders et al. (2000) / Europe	complete products retail	redistribution	same product	customer	third party logistics service provider	Mail Order Company, Wehkamp	Same market	Wehkamp	commercial return (reimburse-ment)	economic (to attract and keep customers)
Swinkels and Van Esch (1998) / Europe	Beer kegs	redistribution / re-use	same product	restaurants, bars, etc.	Bavaria, Agents	Bavaria	Bavaria	Bavaria	functional returns (distribution item)	economic (cost savings)
Toktay et al. (2000) / North-America	single-use photo cameras	remanu-facturing	same product	consumer	Photo Shops / Retailers	Kodak	Same chain	Kodak	end-of-use (or distribution item)	economic (cost savings)
Van der Laan (1997) / Europe	used car parts	remanu-facturing	same product	national importer organizations	Volkswagen (Kassel)	Volkswagen (Kassel)	national importer organizations	Volkswagen (Kassel)	end-of-use	economic

provided to people with handicaps. They assess how many are needed to meet all demands.

End-of-life returns cases

Fleischmann (2000) describes the dismantling of returned, end-of-life computers into useable spare parts with IBM. This case study shows how return obligations can be used as a cheap source for spare parts for systems on which one does not want to spend too much. The problems identified were a lack of knowledge of what actually was in the returned computers as well as an insufficient information system to handle the operations.

Klausner and Hendrickson (2000) develop a model to determine the optimal buy-back amount to guarantee a continuous flow of remanufactured power-tools. The authors apply the model to the actual voluntary take-back program in Germany, where costs go beyond profits.

6.2 Remarks

We have grouped the presentation of the cases according to the return reason as follows: commercial returns, service returns, end-of-use and finally end-of-life returns. This seems a natural way of grouping and discriminating the reverse logistics issues rising from each inventory system. Other authors have done the same (see Dekker and Van der Laan, 2003). It remains to investigate in which degree inventory systems' characteristics are really dependent of the return reason, as well as on the type of product *(what)*.

Many have defended that product data are essential for efficient handling of returns. For instance, Kokkinaki et al. (2003) provide an example of the value of information for disassembly. Other authors have investigated the impact of data on the movement of return and quantity influence inventory management performance, including (Kelle and Silver, 1989; Toktay et al., 2000; De Brito and Van der Laan, 2003). Yet, there is room to model the impact of having a priori information on what can be recovered, i.e. on which parts are likely to be recoverable. In practice, the existent forecasting techniques would have to be enriched with broader explanatory variables. We refer to Toktay (2003) for a discussion of other factors influencing returns, which are potential explanatory variables in advanced forecasting models.

Many authors have investigated the impact of data on inventory management performance in the context of reverse logistics, including (Kelle and Silver, 1989; Toktay et al., 2000; De Brito and Van der Laan, 2002). Yet, there is room to model the impact of having a priori information on what can be recovered, i.e. on which parts are likely to be recoverable. In practice, the existent forecasting techniques would have to be enriched with broader explanatory variables. We refer to (Toktay, 2003) for a discussion of other factors influencing returns, which are potential explanatory variables in advanced forecasting models.

7 Case Studies on Planning and Control of Reverse Logistics Activities

This section deals with the planning and control of reverse logistics activities. The planning and control of reverse logistics activities are strongly related to inventory management, the topic dealt with in the previous section. The latter includes the inventory levels that trigger the execution of activities. We are left with the decisions on lot sizes and scheduling. The case studies that we found can be divided in case studies primarily dealing with collection, disassembly, repair and reassembly, and case studies where attention is paid to the combined planning and control of distribution and collection, and production and processing. The general data on the following case studies is presented in Table 5.

7.1 The case studies

Collection

In a number of case studies, the lot size used for collection is just given, without explaining how this lot size has been determined. All the case studies that we found concern end-of-use or end-of-life returns.

Andriesse (1999) describes the Packaging Return System of Philip Morris Holland BV for among others reusable pallets. Philip Morris Holland BV and most of its suppliers agreed that the lot size for returning empty pallets should be a full truck load. Del Castillo and Cochran (1996) describe the mathematical model used by EMSA, a producer of soft drinks in Mexico City, for determining the quantities of refillable bottles to be returned to the bottling plants from the final customers for the soft drink, via the stores selling the soft drink and the depots delivering the soft drinks to the stores. Duhaime et al. (2001) present the mathematical model used by Canada Post to determine the number of empty containers that should be distributed and returned each month, as well as the number of containers stored per region each month. Klausner and Hendrickson (2000) mention the lot size used for the collection of power tools by Robert Bosch GmbH. Bartels (1998), describing the Dutch nationwide network for collecting and processing portable batteries, pays among others attention to the collection of batteries at municipality collection points. These points can call one of the contracted collectors once a month to collect. Van Donk (1999) describes the system setup by Nelis Utitlieitsbouw BV, a Dutch building company, to keep the rest flows of different types of materials separated at building locations in order to allow a higher level of reuse of these flows, resulting in lower costs. Among others, attention is paid to the number and sizes of the containers used for collection. Whenever a container at a building location is expected to be completely filled soon, a recycling company is called to replace the filled container by an empty one. Van Notten (2000), discussing the bring and pickup systems for the collection of glass from households in the Netherlands, also pays attention to the sizes of the containers used and the collection scheme, often once

Table 5. Case studies focussing on Planning and Control

Reference / Geographical area	Product -in	Process	Product-out	Supplier (owner / user)	Collector	Processor	Customer	Initiator	Reason return Sender (owner / user)	Driver initiator
Andriesse (1999) / Europe	pallets	redistribu-tion	pallets	suppliers Philip Morris Holland	Philip Morris Holland	Philip Morris Holland	suppliers Philip Morris Holland	Philip Morris Holland	cost reduction	economic
Bakkers and Ploos van Amstel (2000) / Europe	PVC lamellas	recycling	PVC material	households companies	distributors lamellas	Ortes Lecluyse	Ortes Lecluyse	Ortes Lecluyse	end-of-life (plus no disposal costs)	economic (green image)
Bartels(1998) / Europe	See Table 2 Network Structure									
Bentley et al. (1986) / North-America	subway cars / transit cars	repair, refurbis-hing	repaired, refurbished cars	NY City et al.		Morrison-Knudsen Company	NY City et al.	Morrison-Knudsen Company		economic
Del Castillo and Cochran (1996) / Central-America	See Table 2 Network Structures									
Duhaime et al. (2001) / North-America	See Table 2 Network Structures									
Driesch et al. (1998) / Europe	See Table 3 Relationships									
Guide and Spencer (1997), Guide et al (1997), Guide and Srivastava (1998) / Noth-America	aircraft, engines, avionics equip-ment	remanu-facturing	remanu-factured aircraft, engines, avionics equipment	US Air Force	US Air Force	US Air Force	US Air Force	US Air Force		

Table 5 continued

Gupta and Chakraborty (1994) / Asia	scrap production glass	recycling	materials for glass production	producer glass	producer glass	producer glass	producer glass	producer glass	costs savings	economic (costs savings)
Klausner and Hendrickson (2000) / Europe	See Table 4 Inventory Management									
Krikke et al. (1999b) / Europe	PC monitors	parts retrieval, recycling	parts, materials	PC owners / users	Roteb (municipal waste company)		secondary markets	Roteb (municipal waste company)		legislation, economic
Robinson (1992) / North-America	diesel engine components	remanu-facturing	remanu-factured diesel engine component			Detroit Diesel remanu-facturer West		Detroit Diesel remanufacturer West		
Schinkel (2000) / Europe	gypsum	recycling	gypsum	building companies	third party LSP	producers gypsum products	customers producers gypsum products	Organisation of producers gypsum products	end-of-life	legislation (lower disposal costs)
Simons (1998) / Europe	parts of sheets (external), production scrap (internal)	recycling	materials	building companies (external), Trespa (internal)	third party LSP, building companies, demolishers	Trespa	Builders	Trespa	end-of-use, end-of-life (reduced disposal costs)	economic, (expected) legislation
Simons (1998) / Europe	pallets	redistri-bution	pallets	builders, demolishers	third party LSP, building companies, demolishers	Trespa	Trespa	Trespa		economic, legislation
Spengler et al. (1997) / Europe	See Table 2 Network Structures									

Table 5 continued

Teunter et al. (2000) / Europe	by-products	re-use	materials	Schering AG	Schering AG	Schering AG	Schering AG	Schering AG	manufactu-ring returns (by-product)	economic
‘t Slot and Ploos van Amstel (1999) / Europe	white and brown goods	recycling	materials	households	municipali-ties, retailers	recycler	users materials	government	end-of-life (no disposal costs)	Legislation, economic,
Thomas Jr (1997) / North-America	airplane engines	remanu-facturing	airplane engines				Pratt & Whitney aircraft West Virginia	Pratt & Whitney aircraft West Virginia		
Tucker et al. (2000) / Europe	See Table 3 Relationships									
Ubbens (2000) / Europe	metal cans for drinks, food	recycling	materials	households etc	municipali-ties (via household waste)	recyclers	metal industries	Stichting Blik		legislation
Van Donk (1999) / Europe	leftovers from building activities	recycling		buidling companies	special collectors				manufactu-ring leftovers (plus lower disposal costs)	responsi-bility, legislation
Van Notten (2000) / Europe	See Table 2 Network Structures									
Wijshof (1997) / Europe	See Table 2 Network Structures									

a week. Schinkel (2000) describes the Dutch nation wide system for the collection of gypsum. Attention is paid to the actual collection of gypsum rest flows via special containers and bags. 't Slot and Ploos van Amstel (1999) describe the pilot project in and around Eindhoven, the Netherlands, preceeding the introduction of a nationwide network for the collection and processing of disposed white and brown goods. Examples of white goods are refrigerators, washing machines, whereas examples of brown goods are PC's, TV and radio sets. Among others attention is paid to the collection scheme at households being fixed (once a month or quarter), and the collection frequency at selling points of white and brown goods. Ubbens (2000), describing the recovery of metal from metal packaging in the Netherlands, pays among others attention to the number and sizes of special containers for the collection from households. Wijshof (1997) describing the collection and processing of the rockwool produced by Rockwool Benelux in the Netherlands, pays among others attention to the sizes that are used for collection and the number of bags that has to be filled before a third party is collecting them for Rockwool after having been contacted by the disposer. Finally, Tucker et al. (2000) pay attention to the consequences of changing the kerbsite collection frequency for newspapers at households for recycling from once per two weeks to once per four weeks in the Borough of Fylde in Lancashire in Great Britain, resulting in a very small decrease in the total quantity collected.

Processing

Bentley et al. (1986) mention that Morrison-Knudsen uses MRP II to plan the remanufacturing of subway overhaul, but the authors do not explain how. This also holds for Robinson (1992), who mentions the use of MRP by Detroit Diesel Remanufacturing West, where Detroit Diesel engines are remanufactured. Driesch et al. (1997), describing the recovery network for engines of Mercedes-Benz cars and vans in Europe, also describe the planning and control of the actual processing of collected engines in the central recovery plant in Berlin, Germany. Among others it is mentioned that the disassembly, cleaning, test, remanufacturing and reassembly activities are dealt with in lots, and that the number of engines that are disassembled is related to the number of reconditioned engines that are reassembled. Aso here no further details are given, nor is explained how these lot sizes have been determined. Guide and Srivastava (1997a) discuss the application of the drum-buffer-rope principle in a US Navy Overhaul/repair Depot consisting of a disassembly, remanufacturing and a reassembly shop. Guide and Spencer (1997) and Guide et al. (1997) discuss a method for rough cut capacity planning for the above depot.

Thomas Jr (1997) mentions that the Pratt Whitney Aircraft remanufacturing facility in West Virginia uses MRP to schedule inspection and rebuild of military and commercial aircraft engines. The batch size is one because different engines have to go through different routings. Bottleneck is the engine reassembly. Buffer time is used to protect this activity form variations in foregoing activities. This time is determined via LP, but no formulas are given. Spengler et al. (1997) discuss an MILP model used for the planning of processing components arising from the dismantling of buildings in the Upper Rhine Valley in Germany.

Combined collection and distribution

Simons (1998) describes the system setup by Trespa International B.V. This company produces sheets made from resins and wood fibers which are used in the building industry. The recycling system was setup to recycle or incinerate (parts of) these sheets. Among others attention is paid to the collection of (parts of) the sheets leftover from building activities. These leftovers are put into containers supplied by Trespa. The customers let Trespa know when a container is filled. Filled containers are replaced by empty containers when new Trespa sheets are delivered to the customers. The reusable pallets used for the distribution of the Trespa sheets to the customers are collected either by third party logistics service providers who get money for the pallets that they deliver to Trespa, or by third party logistics service providers appointed by Trespa for the above collection. In the latter case, at least 15 pallets should be available at the moment of collection. Trespa promises its customers to collect these pallets within 3 weeks. The third party logistics service providers working for Trespa pickup the reusable pallets when they deliver new Trespa sheets.

Bakkers and Ploos van Amstel Jr (2000) describe the system setup by Ortes Lecluyse, a Dutch producer of PVC lamellas, for the recycling of these lamellas. Thereby also attention is paid to the sizes of the containers used for collection and the frequency for emptying these containers located at their direct customers, being once a week when new lamellas are delivered.

Combined processing and production

Gupta and Chakraborty (1984) describe the processing of glass scrap generated during the production of glass. A mathematical model is presented to determine the optimal production lot size, taking into account the recycling activities. Teunter et al. (2000) describe the mathematical model used by Schering AG, a German producer of medicines, for the planning of their production activities, including the processing of the by-products resulting from the production of medicines.

7.2 Remarks

In many of the case studies we found, only some planning and control issues are globally described, most of the time being the lot sizes that are used, without any further explanation. The authors hardly found descriptions of the planning and control concept for reverse logistics, let alone the quantitative motivation behind it. The case studies found hardly give insight into the problems companies have with the planning and control of their reverse logistics activities, nor in the results obtained with their planning and control concept.

On the other hand, quite a number of planning and control concepts for product recovery have been presented in academic literature, many of them including mathematical models to estimate the usefulness of the concepts. For an overview, see (Dekker et al., 2004).

Many of these concepts assume autonomous supply of products that might be recovered, i.e., no direct, explicit relation between the number of products sold or leased, and the number of products returned is assumed, which is one of the essential differences between reverse logistics and many other production situations. Exceptions are literature on repair, see (Guide and Srivastava, 1997b) for an overview, and (Guide and Van Wassenhove, 2001; Guide et al., 2001). Uncertainty has been incorporated as far as the arrival of products for recovery and the duration of repair related activities are concerned. Uncertainty with respect to the result of the processing activities is hardly ever taken into account. An exception is (Souza et al., 2002).

Concluding, it seems useful to do more case studies to estimate which planning and control concepts are used in practice, how the values of the different parameters related to them are calculated, and how well they are performing. It also seems worthwhile to estimate the usefulness of the theoretical concepts developed or under development.

8 Case Studies on IT for Reverse Logistics

We have found various cases concerning applications of IT for reverse logistics activities (see Table 6). IT is used to support reverse logistics during different phases in the life cycle of a product, namely manufacturing, distribution and market (i.e., use), see also (Kokkinaki et al., 2003; Hendrickson et al., 2003). Table 7 shows more details about the case studies found on IT. Apart from listing the type of IT tool, it also shows which information it requires, the type of support it gives and for which lifecycle phase this is appropriate.

8.1 The case studies

Manufacturing

Regarding the phase of product development and actual manufacturing, there are two variables to consider within the *what* dimension: material content and product structure. The materials that are used and how they are combined determine the degree and the type of a potential recovery once the product is at the end of its life. Marking parts with manufacture identification are also helpful when a product has to be pulled out of the market due to a defect, i.e. product recalls (Smith, 1996). Many companies have already in place product development programs encompassing design for the environment, for recovery, for disassembly, and so on - generally called as Design for X, or just DfX. This holds for instance for Xerox Europe (Maslennikova and Foley, 2000). Xerox has an extensive Design-for-the-Environment program in place, where the design of each new component has to be accompanied with instructions for what to do with them at the end-of-use.

Table 6. Case studies focusing on Information and Technology

Reference / Geographical area	Product-in	Process	Product-out	Supplier (owner / user)	Collector	Processor	Customer	Initiator	Reason return Sender (owner / user)	Driver initiator
Nagel and Meyer (1999) / Europe	refrigera-tors	remanu-facturing, recycling	plastics, metals	end-of-life user	dealer, service provider	recycling plant			end-of-life	legislation, economic
Nagel and Meyer (1999) / Europe	computers	parts retrieval	computer components	end-of-life user	VOBIS (retailer)	Covertronic (recycler)		Covertronic (Recycler)	end-of-life	economic
Meyer (1999) / North-America	cosmetics	resale	cosmetics	customer	Estee Lauder	Estee Lauder	same chain, employees	Esteé Lauder	reimburse-ment	legislation, economic
Farrow et al. (2000) / North-America	post-consumer plastics	recycling	plastics	households, companies	supplier of recycled resins	manufacturer kayak	buyers kayak	Walden Paddlers	end-of-life (waste disposal)	corporate citizenship
Linton and Jonhson (2000) / North-America	circuit boards	remanu-facturing	circuit boards	customer	customer Service (NN)	customer Service (NN)	same chain	Nortel Networks	commercial returns	economic
Klausner et al. (1998, 2000) / Europe	See Table 4 Inventory Management									
Maslennikova and Foley (2000) / Europe	Recycled materials, Xerox products	refurbishing recycling	reprocessed Xerox products	customer	Xerox Europe, Ltd. service engineers	Xerox Europe, Ltd.	same chain	Xerox Europe, Ltd.	end-of-use (no longer needed)	economic, legislation

Table 7. Case studies focusing on IT tools, requirements and type of support for reverse logistics

Reference / Geographical area	IT Tool	Information requirements	Type of support	Life cycle phase
Klausner et al. (1998) , Klausner and Hendrickson (2000) / Europe	EDL for electric motor reuse	potential cost savings	Reuse decisions through information on usage patterns	Product development, Market
Nagel and Meyer (1999) /Europe	DSS for end-of-use	operations costs and recycling revenues	Cost optimization, facilities location, vehicle routing, etc.	(Re-) distribution
Nagel and Meyer (1999) / Europe	Computer configuration reader	operations costs and recycling revenues	Setting buy-back price	Market
Meyer (1999) / North-America	software for return handling	product's expiration date, damage check	Recovery-related decisions	(Re-) distribution
Linton and Jonhson (2000)/ North-America	DSS for remanufacturing	processing data	Remanufacturing-related decisions	(Re-) distribution
Maslennikova and Foley (2000) / Europe	DfX, remote maintenance, etc.	product data	Recovery options, environmental sustainability, etc.	All phases
Farrow et al. (2000)/ North-America	DfX (X= Recyclability)	separation of resins, technological innovation	Design, production, recycling	(Re-) distribution

Recovery can also be the starting point for product development, as it is the case of Walden Paddlers, who launched a 100% recycled kayak project (Farrow et al., 2000). The project had to rely much on computer experiments as no design then available suited recycled resins. The company was able to attract a manufacturer to invest in advanced rotational molding technology and to convince the supplier to proceed to further resins' separation.

Distribution

Landers et al. (2000) highlight the importance of tracking component's orders in the case of a closed-loop business telephones supply chain. The authors use a concept called "virtual warehousing" where real-time information feeds expeditious algorithms to support decisions. The use of ICT leads to an improvement in stock levels, routing and picking processes when compared with the pre-ICT scenario. Xerox uses bar code labels to track packaging material with the aim of achieving resources' preservation (Maslennikova and Foley, 2000).

The Fraunhofer IML institute has developed software to embed data on recovery processes as reported in (Nagel and Meyer, 1999). The authors consider two national reverse logistics networks in Germany: one for the recycling of refrigerators and one for the recycling of computers. Costs could be minimized by optimization of the location of facilities, vehicle routing and operations' scheduling supported by the software. For the case of the German computer-recycling network, transport volume (in tons per km) could be reduced by almost 20%.

Market

After the customer has accepted the product and starts using it, the product may need maintenance. Xerox has a remote faulty detection system in place called the Sixth Sense (Maslennikova and Foley, 2000). In some situations, the problem is identified at a distance and solved. Customers are assisted by a multi-functional database that permits them to get thorough acquaintance with product characteristics.

In the case of Nortel Networks, a Decision Support System (DSS) was developed to assist remanufacturing (Linton and Jonhson, 2000). The tool permits to apprehend the interrelations between the production and the remanufacturing of products. By the DSS both processes can be better planned and controlled, resulting in a more efficient allocation of resources.

Estee Lauder is another firm that has developed specialized software to handle product returns (Meyer, 1999). The system checks the returned cosmetics with respect to their expiration date and damages. In this way, recovery related decisions are accelerated. The software is linked to an automatic sorting system, which saves labor costs. Estee Lauder could reclaim the investment on ICT within one year's time. Nagel and Meyer (1999) report on software developed by the German recycler Covertronic to read the configuration of a computer and to compute costs and revenues of subsequent recovery. Based on this, an appropriate bonus is offered to the supplier when the computer is returned. Covertronic operates this software together with Vobis, a large computer retailer in Germany. Another technology available is the so-called electronic data logger (EDL). This device is able to store data on physical parameters, which can be retrieved later. The idea is to put them into products or equipment (as is done for some coffee machines) and to register information about heat or other parameters as they are used. Thus, at the point of recovery, one could make use of this information to decide which destiny to give to certain product without first investing resources in disassembling and testing components. Klausner et al. (1998) have investigated the benefits of collecting information via this chip technology in power tools, whereas Simon et al. (2001) apply both steady state and transient models to evaluate the benefits of using a data logger.

8.2 Remarks

Table 7 summarizes the IT tools to support reverse logistics activities described in the case studies presented above. The case studies illustrate IT applications in all the phases of the life cycle of a product and show how they may contribute to improvements in reverse logistics. For instance, Xerox (Maslennikova and Foley, 2000) has an integrated solution for reverse logistics from product development to recovery or proper disposal. All the case studies we found, provide insight in the benefits of IT. However, IT tools are very demanding regarding data on reverse logistic processes and associated costs and earnings. Often, this data is not available. Nagel and Meyer (1999) state that the lack of data is a bottleneck, which complicates the management of recycling systems. The DSS of Nortel Networks could not be designed as desired due to a shortage of data on returns from customers (Linton and Johnson, 2000). Besides this, to acquire and to manage the data is very expensive. However, the required investments in technologies or otherwise gathering data are not reported, except for the Estee Lauder case (Meyer, 1999) and partially for the Walden Paddlers case (Farrow et al., 2000). Therefore, the real benefit of IT investment is difficult to assess. In face of limited investment capacity, it would be helpful to know in which phase the investment would contribute most to the earnings of a company. In addition, if alternative technologies are available, one could investigate which one is the best. To do so, one has to take into account: costs and benefits of collecting and managing data and costs of investing and managing the technology. The following references are an excellent starting point, because they analysis the potential advantages or bottlenecks of some type of information in the context of reverse logistics: (Kelle and Silver, 1989; Inderfurth and Jensen, 1999; Klausner et al., 1998; Simon et al., 2001; De Brito and Van der Laan, 2003; Ferrer and Ketzenberg, 2003; Toktay, 2003).

9 Summary and Conclusions

In this review we provided a content analysis of more than 60 cases of reverse logistics practices. With this content analysis, 1) we gave a substantial overview of the diversity of real life reverse logistics situations; and 2) we provided a reference guide to researchers searching for case support. The analysis was based on the framework for Reverse Logistics recently proposed by De Brito and Dekker (2004). Thus, for each case study we answered the following questions: *what* products/materials are being returned? *how* are they being recovered? *who* are the parties involved? and *why* is all this happening?

We presented the case studies according the following decision-making focus: Network Structures, Relationships, Inventory Management, and Planning and Control. Furthermore, there was an overview of Information and Technology (IT) for reverse logistics, based on the case studies.

The following observations come forward from this review.

- The how and who typologies can be used to characterize reverse logistic networks.
- The usual incentive tools in forward logistics are also employed in reverse logistics, which has only one extra tool, viz. deposit fees. There is a lack of literature supporting the *choice of incentive.*
- The *why-supplier* typology seems a natural way of grouping and discriminating the reverse logistics issues rising from the Inventory Management cases.
- Case studies on Planning and Control of product recovery were found on
  - separate collection of (parts of) products for recovery
  - separate processing of (parts of) products for reuse or disposal
  - combined planning and control of collection of products for recovery and distribution of new products:
  - combined planning and control of processing products for recovery and production of new products
- Cases on the use of IT exist for all the stages of the life-path of a product (product development, supply chain, and use with customer) with benefits for reverse logistics.
- Though the technology to process and transmit information useful for reverse logistics seems to be available, the lack of appropriate data is still a bottleneck in the implementation of reverse logistic decision support systems.

Acknowledgements

The research presented in this paper was supported by the European Commission as part of the TMR project REVLOG (ERB 4061 PL 97-650), the European Working group on Reverse Logistics (http://www.fbk.eur.nl/OZ/REVLOG/). The first author would also like to acknowledge the financial support given by the Portuguese Foundation for the Development of Science and Technology, "Fundação para a Ciência e a Tecnologia." We also would like to thank Moritz Fleischmann for his comments on the Network Structures section.

References

Andriesse, F. G. (1999): Successful implementation of reverse logistics at Philip Morris (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Anupindi, R. / Bassok, Y. (1999): Supply Contracts with Quantity Commitments and Stochastic Demand. In: Tayur, S. / Ganeshan, R. / Magazine, M. (eds.): Quantitative models for supply chain management. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 197-232

Bakkers, F. T. P. G. / Ploos van Amstel Jr, J. J. A. (2000): Ortes Lecluyse return- and recycling system for PVC lamellas (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. /

Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Barros, A. I. / Dekker, R. / Scholten, V. (1998): A two-level network for recycling sand: A case study. *European Journal of Operational Research* 110(2):199-215

Bartel, T. (1995): Recycling Program for Toner Cartridges and Optical Photoconductors. In: Proceedings IEEE Symposium on Electronics and the Environment, Orlando Florida, May 1-3, pp 225-228

Bartels, J. J. C. (1998): Collection and processing of batteries in The Netherlands - the Stibat execution scheme (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Bentley, D. A. / Rothstein, S. R. / Witt, R. L. (1986): Applying MRP II to Remanufacturing: A Case Study. In: APICS Conference Proceedings pp 18-21

Chang, N-B. / Wei, Y. L. (2000): Siting recycling drop-off stations in urban area by genetic algorithm-based fuzzy multi-objective nonlinear integer programming modeling. Fuzzy Sets and Systems 114:133-149

Corbett, C. J. / Tang, C. S. (1999): Designing Supply Contracts: Contract Type and Information Asymmetry. In: Tayur, S. / Ganeshan, R. / Magazine, M. (eds.): Quantitative models for supply chain management. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 269-298

De Brito, M. P. / Dekker, R. (2003): Modeling product returns in inventory control - an empirical validation of general assumptions. *International Journal of Production Economics* 81-82: 225-241

De Brito, M. P. / Dekker, R. (2004): A framework for Reverse Logistics. In: Dekker, R. / Inderfurth, K. / van Wassenhove, L. N. / Fleischmann, M. (eds.): Reverse Logistics: Quantitative Models for Closed-Loop Supply Chains, Springer-Verlag, Berlin Heidelberg New York, pp 1 – 20

De Brito, M. P. / Laan, E. A. van der (2002): Inventory control with product returns: the impact of (mis)information. Econometric Institute Report EI 2002-29, Erasmus University Rotterdam, Rotterdam, The Netherlands

Del Castillo, E. / Cochran, J. K. (1996): Optimal short horizon distribution operations in reusable container systems. *Journal of the Operational Research Society* 47(1):48-60

De Koster, M. B. M. / Flapper, S. D. P. / Krikke, H. R. / Vermeulen, W. S. (2000): Reverse Logistics in the wholesale white-good sector (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

De Koster, M. B. M. / Vendel, M. A. van de / De Brito, M. P. (2001): How to organise return handling: an exploratory study with nine retailer warehouses. *International Journal of Retail & Distribution Management* 30(8):407-421

Dekker, R. / Inderfurth, K. / van Wassenhove, L. N. / Fleischmann, M. (eds.) (2004): Reverse Logistics: Quantitative Models for Closed-Loop Supply Chains. Springer Verlag, Berlin Heidelberg New York

Diaz, A. / Fu, M. C. (1997): Models for multi-echelon repairable item inventory systems with limited repair capacity. European Journal of Operational Research 97(3):480-492

Dijkhuizen, H. P. (1997): Reverse Logistics at IBM (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Donker, S. / Ploeg, T. van der (2001): Multi-echelon inventory control of repairable spare parts at Lucent Technologies Nederland (in Dutch). In: van Goor, A. R. / Flapper, S.

D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Driesch, H. M. / Flapper, S. D. P. / Oyen, J. E. van (1998): Logistic control of engine reuse at Daimler-Benz MTR (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Duhaime, R. / Riopel, D. / Langevin, A. (2001): Value analysis and optimization of reusable containers at Canada Post. *Interfaces* 31(3): 3-15

Faria de Almeida, A. / Robertson, A. (1995): Domestic Batteries and the Environment: A Life-Cycle Approach to Consumer Electronic Products. In: Concept Proceedings International Conference Clean Electronics Products & Technology, Edinburgh, Scotland, pp 162-167

Farrow, P. H. / Johnson, R. R. / Larson, A. L. (2000): Entrepreneurship, Innovation and Sustainability Strategies at Walden Paddlers Inc., *Interfaces* 30(3):215-225

Ferrer, G. / Ketzenberg, M. (2003): Value of yield information for remanufacturing complex products. Working paper, Kenan-Flagler Business School, University of North Carolina at Chapel Hill

Flapper, S. D. P. / van Nunen, J. A. E. E. / Van Wassenhove, L. N. (eds.) (2004): Managerial aspects of closed loop supply chains. Springer-Verlag, Heidelberg

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. Lecture Notes. *Economics and Mathematical Systems*, Volume 501. Springer-Verlag, Berlin

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

Fleischmann, M. / Beullens, P. / Bloemhof-Ruwaard, J. M. / Van Wassenhove, L. N. (2001): The impact of product recovery on logistics network design. *Production and Operations Management* 10(2):156-173

Fleischmann, M. / Beullens, P. / Bloemhof-Ruwaard, J. M. / Beullens, P. / Dekker, R. (2004): Reverse logistic network design. In: Dekker, R. / Inderfurth, K. / Van Wassenhove, L. N. / Fleischmann, M. (eds.): Reverse Logistics: Quantitative Models for Closed-Loop Supply Chains, Springer-Verlag, Berlin Heidelberg New York, pp 1 – 20

Ganeshan, R. / Jack, E. / Magazine, M. J. / Stephens, P. (1999): A taxonomic review of supply chain management research. In: Tayur, S. / Ganeshan, R. / Magazine, M. (eds.): Quantitative models for supply chain management. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 840-879

Goggin, K. / Browne, J. (2000): Towards a taxonomy of resource recovery from end-of-life products. *Computers in Industry* 42:177-191.

Guide Jr, V. D. R. (2000): Production planning and control for remanufacturing: industry practice and research needs. *Journal of Operations Management* 18:467-483

Guide Jr, V. D. R. / Spencer, M. S. (1997): Rough-cut capacity planning for remanufacturing firms. *Production Planning & Control* 8(3):237-244

Guide Jr, V. D. R. / Srivastava, R. (1997a): Buffering from material recovery uncertainty in a recoverable manufacturing environment. *Journal of the Operational Research Society* 48:519-529

Guide Jr, V. D. R. / Srivastava, R. (1997b): Repairable inventory theory: models and applications. *European Journal of Operational Research* 102:1-20

Guide Jr, V. D. R. / Srivastava, R. (1998): Inventory buffers in recoverable manufacturing. *Journal of Operations Management* 16:551-568

Guide Jr, V. D. R. / Srivastava, R. / Spencer, M. S. (1997): An evaluation of capacity planning techniques in a remanufacturing environment. *International Journal of Production Research* 35:67-82

Guide Jr, V. D. R. / Teunter, R. H. / Van Wassenhove, L. N. (2001): Matching Supply and Demand to Maximize Profits from Remanufacturing. Working Paper 2001/84/TM/CIMSO 21, INSEAD, Fontainebleau Cedex, France

Guide Jr, V. D. R. / Van Wassenhove, L. N. (2001): Managing product returns for remanufacturing. *Production and Operations Management* 10(2):142-155

Gupta, T. / Chakraborty, S. (1984): Looping in a multistage production system. *International Journal of Production Research* 22:299-311

Hendrickson, C. / Scott Matthews, H. / Cagan, J. / McMichael, F. C. (2003): Green Engineering. In: Guide Jr, V. D. R. / van Wassenhove, L. N. (eds.): Business Perspectives in Closed-Loop Supply Chains, Carnegie Mellon University Press, Pittsburgh, pp 291-314

Inderfurth, K. / Jensen, T. (1999): Analysis of MRP Policies with Recovery Options. In: Leopold-Wildburger, U. / Feichtinger, G. / Kistner, K. P. (eds.): Modelling and Decisions in Economics, Physica-Verlag, Heidelberg-New York, pp 189-228

Kelle, P. / Silver, E. A. (1989): Forecasting the Returns of Reusable Containers. *Journal of Operations Management* 8(1):17-35

Klausner, M. / Grimm, W. M. / Hendrickson, C. (1998): Reuse of electric motors in consumer products, design and analysis of an electronic data log. *Journal of Industrial Ecology* 2(2):89-102

Klausner, M. / Hendrickson, C. (2000): Reverse-Logistics Strategy for Product Take-Back. *Interfaces* 30(3):156-165

Kleineidam, U. / Lambert, A. J. D. / Blansjaar, J. / Kok, J. J. / Heijningen, R. J. J. (2000): Optimising product recycling chains by control theory. *International Journal of Production Economics* 66(2):185-195

Kokkinaki, A. / Zuidwijk, R. / Nunen, J. van / Dekker, R. (2003): Information Management for Reverse Logistics. In: Dekker, R. / Inderfurth, K. / van Wassenhove, L. N. / Fleischmann, M. (eds.): Reverse Logistics: Quantitative Models for Closed-Loop Supply Chains, Springer-Verlag, Berlin Heidelberg New York, pp 381-405

Kopicki, R. J. / Berg, M. J. / Legg, L. / Dasappa, V. / Maggioni, C. (1993): Reuse and Recycling: Reverse Logistics Opportunities. Council of Logistics Management, Oak Brook, IL

Kostecki, M. (1998): The durable use of consumer products: new options for business and consumption. Kluwer Academic Publishers, Dordrecht, The Netherlands

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

Krikke, H. R. / Harten, A. van / Schuur, P. C. (1999b): Business case Roteb: recovery strategies for monitors. *Computers & Industrial Engineering*, 36(4):739:757

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

Landers, T. L. / Cole, M. H. / Walker, B. / Kirk, R. W. (2000): The virtual warehousing concept. Transportation Research Part E, *Logistics & Transportation Review* 36E(2): 115-125

Lariviere, M. A. (1999): Supply Chain Contracting and Coordination with Stochastic Demand. In: Tayur, S. / Ganeshan, R. / Magazine, M. (eds.): Quantitative models for supply chain management. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 233-268

Linton, J. D. / Jonhson, D. A. (2000): A decision support system for planning remanufacturing at Nortel Networks. *Interfaces* 30: (6) 17-31

Louwers, D. / Kip, B. J. / Peters, E. / Souren, F. / Flapper, S. D. P. (1999): A facility location allocation model for reusing carpet materials. *Computers & Industrial Engineering* 36:855-869

Lund, H. F. (2001): The McGraw-Hill recycling handbook. McGraw-Hill, New York

Maslennikova, I. / Foley, D. (2000): Xerox's approach to sustainability. *Interfaces* 30(3):226-233

McGavis, D. (1994): The energy bucket and a not-so-drop-in-the-bucket portion of waste stream, consumables. In: IEEE International Symposium on Electronics & the Environment, San Francisco, California, May 2-4, pp 267-272

Meijer, H. W. (1998): Canon Benelux and Reverse Logistics (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Meyer, H. (1999): Many happy returns. *The Journal of Business Strategy* 20(4): 27-31

Moffat, J. (1992): Resources for the repair of non-modular aero-engines. *Omega* 20(1): 99-103

Nagel, C. / Meyer, P. (1999): Caught between ecology and economy: end-of-life aspects of environmentally conscious manufacturing. *Computers & Industrial Engineering* 36(4): 781-792

Realff, M. J. / Ammons, J. C. / Newton, D. (2000): Strategic design of reverse production systems. *Computers and Chemical Engineering* 24: 991-996

Robinson, S. J. (1992): The World of a Class "A" Remanufacturing Company. In: APICS Remanufacturing Seminar Proceedings pp 1-3

Rogers, D. S. / Tibben-Lembke, R. S. (1999): Going Backwards: reverse logistics trends and practices. Reverse Logistics Executive Council, Pittsburgh, PA

Rudi, N. / Pyke, D. F. / Sporsheim, P. O. (2000): Product recovery at the Norwegian National Insurance Administration. *Interfaces* 30(3):166-179

Sanders, H. / Harten, A. van / Heijden, M. C. van der (2000): Logistics and reverse logistics at Wehkamp (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Schinkel, H. (2000): Plaster remnants should not be thrown away! (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Simons, P. H. W. (1998): Reverse Logistics at Trespa International B.V (in Dutch). In Handboek Reverse Logistics edited by van Goor, A. R. / Flapper, S. D. P. / Clement, C. / Kluwer, B. V., Deventer, The Netherlands

Simon, M. / Bee, G. / Moore, P. / Pu, J-S. / Xie, C. (2001): Modelling of the life cycle of products with data acquisition features. *Computers in Industry* 45: 111-122

Souza, G. C. / Ketzenberg, M. / Guide Jr, V. D. R. (2002): Capacitated Remanufacturing with Service Level Constraints. *Production and Operations Management*, 11(2): 231–248

Spengler, T. / Puchert, H. / Penkuhn, T. / Rentz, O. (1997): Environmental integrated production and recycling management. *European Journal of Operational Research*, 97(2): 308-326

Stake, R. E. (1995): The art of case study. SAGE publications, California, USA

Swinkels, J. / Esch, C. F. A. van (1998): Return packaging and turnover rates (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Teunter, R. / Inderfurth, K. / Minner, S. / Kleber, R. (2000): Reverse logistics in a pharmaceutical company: a case. Working Paper, Otto von Guericke-University of Magdeburg, Magdeburg, Germany

Thomas Jr, M. (1997): Emerging technology: production scheduling matures. *IIE Solutions* January: 25-29

Toktay, L. B. (2003): Forecasting Product Returns. In: Guide Jr, D. V. R. / van Wassenhove, L. N. (eds.): Business Perspectives in Closed-Loop Supply Chains, Carnegie Mellon Press, Pittsburgh, pp 203-219

Toktay, L. B. / Wein, L. M. / Zenios, S. A. (2000): Inventory Management of Remanufacturable Products. *Management Science* 46(11):1412-1426

Tsay, A. A. (2001): Managing retail channel overstock: markdown money and return policies. *Journal of Retailing* 77(4): 457-492

Tsay, A. A. / Nahmias, S. / Agrawal, N. (1999): Modeling Supply Chain Contracts: A Review In: Tayur, S. / Ganeshan, R. / Magazine, M. (eds.): Quantitative models for supply chain management. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 299-336

't Slot, W. M. E. / Ploos van Amstel, J. J. A. (1999): Lessons learnt from the Dutch pilot project Apparetour (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Tucker, P. / Speirs, D. / Smith, D. (2000): The impact of a change in collection frequency on kerbside recycling behaviours. *Journal of Environmental Planning and Management* 43(3): 335-350

Ubbens, U. M. P. (2000): The closed loop of metal packaging (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Van Burik, A. M. C. (1998): Return flow problems of car wreck materials (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Van der Laan, E. A. (1997): The effects of remanufacturing on inventory control. PhD thesis, Erasmus University Rotterdam, Rotterdam, The Netherlands

Van Donk, F. (1999): Environmental logistics in housebuilding and public construction (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Van Notten, R. A. P. (2000): Collection and processing for the re-use of packaging glass (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Vroom, J. A. / Linden, A. van der / Kraal, A. C. (2001): Packaging and Reverse Logistics at Campina Melkunie (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Wijshof, C. P. H. (1997): Reverse Logistics, a case (in Dutch). In: van Goor, A. R. / Flapper, S. D. P. / Clement, C. (eds.): Handboek Reverse Logistics. Kluwer B.V., Deventer, The Netherlands

Yender, G. L. (1998): Battery recycling technology & collection processes. In: Proceedings IEEE International Symposium on Electronics and the Environment, Oak Brook, IL, USA, May 4-6, pp 30-35

Electronic libraries

ABI/Inform, ABI/Inform Global http://www.proquest.umi.com/
APICS: The American Production and Inventory Control Society http://www.apicsorg/
IEEE: Institute of Electrical and Electronics Engineers, Inc, http://www.ieee.org/
NAICS: North American Industry Classification System, http://www.naics.org/
OECD: Organization for Economic Co-operation and Development, http://www.oecd.org/
RevLog: the European Working group on Reverse Logistics http://www.fbk.eur.nl/OZ/REVLOG/
ScienceCitation: Science Citation Index Expanded, Elsevier Science B.V. http://www.sciencedirect.com/

List of Contributors

Enrico **Angelelli**, University of Brescia, Department of Quantitative Methods, C. da S. Chiara 48b, 25122 Brescia, Italy, angele@eco.unibs.it

Claudia **Archetti**, University of Brescia, Department of Quantitative Methods, C. da S. Chiara 48b, 25122 Brescia, Italy, archetti@eco.unibs.it

Angela **Bauer**, Universität Erlangen-Nürnberg, Lehrstuhl für Betriebswirtschaftslehre, insb. Logistik, Theodorstraße 1, D-90489 Nürnberg, Germany, angela.bauer@logistik.uni-erlangen.de

Thomas J. **Bieding**, Bergische Universität Wuppertal, Department of Economics and Business Administration, 42097 Wuppertal, Germany, thomas@bieding.de

Jacqueline M. **Bloemhof**, Erasmus University Rotterdam, Rotterdam School of Management, Faculty of Business Administration, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, j.bloemhof@fbk.eur.nl

Marisa P. **de Brito**, Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, debrito@few.eur.nl

Joachim R. **Daduna**, University of Applied Business Administration at Berlin, Badensche Strasse 50 – 51, D - 10825 Berlin, Germany, daduna@fhw-berlin.de

Rommert **Dekker**, Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, rdekker@few.eur.nl

Simme Douwe P. **Flapper**, Technische Universiteit Eindhoven, Faculty of Technology Management, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, s.d.p.flapper@tm.tue.nl

Roland **Geyer**, University of California, Bren School of Environmental Science and Management, Santa Barbara, USA, geyer@bren.ucsb.edu

Simon **Görtz**, University of Wuppertal, Faculty of Economics and Social Scienes, Gaußstraße 20, 42119 Wuppertal, Germany, simon.goertz@wiwi.uni-wuppertal.de

Andreas **Klose**, University of Zürich, Institute for Operations Research, Moussonstrasse 15, 8044 Zürich, Switzerland, klose@ior.unizh.ch

Herbert **Kopfer**, University of Bremen, Department of Economics and Business Administration, Chair of Logistics, Wilhelm-Herbst-Straße 5, 28359 Bremen, Germany, kopfer@logistik.uni-bremen.de

René (M.) B. M. **de Koster**, Erasmus University Rotterdam, Rotterdam School of Management, Faculteit Bedrijfskunde, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, rkoster@fbk.eur.nl

Harold **Krikke**, CentER Applied Research, affiliated to Tilburg University, P.O. Box 90153, 5000LE, Tilburg, The Netherlands, krikke@uvt.nl

Roelof **Kuik**, Erasmus University Rotterdam, Rotterdam School of Management, Faculty of Business Administration, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, r.kuik@fbk.eur.nl

Tuan **Le Anh**, Erasmus University Rotterdam, Faculteit Bedrijfskunde, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, ltuan@fbk.eur.nl

Tho **Le Duc**, Erasmus University Rotterdam, Rotterdam School of Management, Faculteit Bedrijfskunde, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, tleduc@fbk.eur.nl,

Barbara **Lenz**, German Aerospace Center, Institute of Transport Research, Rutherfordstrasse 2, D-12489 Berlin, Germany, barbara.lenz@dlr.de

Renata **Mansini**, University of Brescia, Department of Electronics for Automation, via Branze 38, 25123 Brescia, Italy, rmansini@ing.unibs.it

Jo A.E.E. **van Nunen**, Erasmus University Rotterdam, Rotterdam School of Management, Faculty of Business Administration, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, j.nunen@fbk.eur.nl

Jörn **Schönberger**, University of Bremen, Department of Economics and Business Administration, Chair of Logistics, Wilhelm-Herbst-Straße 5, 28359 Bremen, Germany, sberger@logistik.uni-bremen.de

Ciska M. **Smeets**, Erasmus University Rotterdam, Rotterdam School of Managment, Faculty of Business Administration, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands,

M. Grazia **Speranza**, University of Brescia, Department of Quantitative Methods, C. da S. Chiara 48b, 25122 Brescia, Italy, speranza@eco.unibs.it

Luk N. **Van Wassenhove**, INSEAD, Boulevard de Constance, 77305 Fontainebleau, France, luk.van-wassenhove@insead.edu

Rob **Zuidwijk**, Erasmus University Rotterdam, Rotterdam School of Management, Faculty of Business Administration, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, rzuidwijk@fbk.eur.nl

Lecture Notes in Economics and Mathematical Systems

For information about Vols. 1–454
please contact your bookseller or Springer-Verlag

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.

Vol. 520: B. Glaser, Efficiency versus Sustainability in Dynamic Decision Making. IX, 252 pages. 2002.

Vol. 521: R. Cowan, N. Jonard (Eds.), Heterogenous Agents, Interactions and Economic Performance. XIV, 339 pages. 2003.

Vol. 522: C. Neff, Corporate Finance, Innovation, and Strategic Competition. IX, 218 pages. 2003.

Vol. 523: W.-B. Zhang, A Theory of Interregional Dynamics. XI, 231 pages. 2003.

Vol. 524: M. Frölich, Programme Evaluation and Treatment Choise. VIII, 191 pages. 2003.

Vol. 525:S. Spinler, Capacity Reservation for Capital-Intensive Technologies. XVI, 139 pages. 2003.

Vol. 526: C. F. Daganzo, A Theory of Supply Chains. VIII, 123 pages. 2003.

Vol. 527: C. E. Metz, Information Dissemination in Currency Crises. XI, 231 pages. 2003.

Vol. 528: R. Stolletz, Performance Analysis and Optimization of Inbound Call Centers. X, 219 pages. 2003.

Vol. 529: W. Krabs, S. W. Pickl, Analysis, Controllability and Optimization of Time-Discrete Systems and Dynamical Games. XII, 187 pages. 2003.

Vol. 530: R. Wapler, Unemployment, Market Structure and Growth. XXVII, 207 pages. 2003.

Vol. 531: M. Gallegati, A. Kirman, M. Marsili (Eds.), The Complex Dynamics of Economic Interaction. XV, 402 pages, 2004.

Vol. 532: K. Marti, Y. Ermoliev, G. Pflug (Eds.), Dynamic Stochastic Optimization. VIII, 336 pages. 2004.

Vol. 533: G. Dudek, Collaborative Planning in Supply Chains. X, 234 pages. 2004.

Vol. 534: M. Runkel, Environmental and Resource Policy for Consumer Durables. X, 197 pages. 2004.

Vol. 535: X. Gandibleux, M. Sevaux, K. Sörensen, V. T'kindt (Eds.), Metaheuristics for Multiobjective Optimisation. IX, 249 pages. 2004.

Vol. 536: R. Brüggemann, Model Reduction Methods for Vector Autoregressive Processes. X, 218 pages. 2004.

Vol. 537: A. Esser, Pricing in (In)Complete Markets. XI, 122 pages, 2004.

Vol. 538: S. Kokot, The Econometrics of Sequential Trade Models. XI, 193 pages. 2004.

Vol. 539: N. Hautsch, Modelling Irregularly Spaced Financial Data. XII, 291 pages. 2004.

Vol. 540: H. Kraft, Optimal Portfolios with Stochastic Interest Rates and Defaultable Assets. X, 173 pages. 2004.

Vol. 541: G.-Y. Chen, Vector Optimization (planned).

Vol. 542: J. Lingens, Union Wage Bargaining and Economic Growth. XIII, 199 pages. 2004.

Vol. 543: C. Benkert, Default Risk in Bond and Credit Derivatives Markets. IX, 135 pages. 2004.

Vol. 544: B. Fleischmann, A. Klose, Distribution Logistics. X, 284 pages. 2004.

Vol. 545: R. Hafner, Stochastic Implied Volatility. XI, 229 pages. 2004.

Vol. 546: D. Quadt, Lot-Sizing and Scheduling for Flexible Flow Lines. XVIII, 227 pages. 2004.