TRANSPORTATION SCIENCE
Vol. 38, No. 2, May 2004, pp. 121–134
ISSN 0041-1655 | EISSN 1526-5447 | 04 | 3802 | 0121

DOI 10.1287/trsc.1030.0061

Economies of Scale in Empty Freight Car Distribution in Scheduled Railways

Martin Joborn
Carmen Consulting, Maria Bangata 6, SE-118 63, Stockholm, Sweden, martin.joborn@carmenconsulting.com

Teodor Gabriel Crainic
Intelligent Transportation Systems Laboratory, Centre de recherche sur les transports, Université de Montréal, C.P. 8888, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7, and Département de management et technologie, École des Sciences de la Gestion, Université du Québec à Montréal, C.P. 6192, Succursale Centre-ville, Montréal, Québec, Canada H3C 4R2, theo@crt.umontreal.ca

Michel Gendreau
Centre de recherche sur les transports, and Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Québec, Canada H3C 3J7, michelg@crt.umontreal.ca

Kaj Holmberg
Department of Mathematics, Linköping Institute of Technology, Linköping University, SE-581 83 Linköping, Sweden, kahol@mai.liu.se

Jan T. Lundgren
Department of Science and Technology, Campus Norrköping, Linköping University, SE-601 74 Norrköping, Sweden, janlu@itn.liu.se

In this paper, we consider empty freight car distribution in a scheduled railway system. We analyze the cost structure for the repositioning of empty cars, and conclude that the distribution cost shows an economy-of-scale behavior. In addition to the cost proportional to the number of cars sent from origin to destination, there is a cost related to car-handling operations at yards, which depends on the number of car groups that are handled. Thus, if we can find a transportation pattern in which fewer but larger groups of cars are built, the total distribution cost can be decreased.

The objective of the paper is to propose an optimization model that explicitly takes this economy-of-scale effect into account. We use a time-dependent network to describe the possible car movements in time and space, and show how this network can be transformed into a network with fixed costs on links representing movements of cars with identical origin and destination terminals. The resulting optimization model is a capacitated network design model, where each capacity constraint limits the flow on several arcs. We describe a tabu heuristic for solving the model, and present computational results.

Key words: empty freight car distribution; time-dependent capacitated network design; tabu search; scheduled railways
History: Received: March 2001; revisions received: December 2002, April 2003; accepted: April 2003.

Introduction

In freight transportation systems, there are often spatial imbalances in the supply and demand of empty transportation equipment such as trailers or freight cars. This is a result of regional imbalances regarding the flows of goods. To compensate for these imbalances, empty equipment has to be *repositioned*, that is, moved from areas with an excess of transportation equipment to areas with a deficit. This paper considers the distribution of empty freight cars in a scheduled railway system. The empty freight car distribution process includes the problem of planning and performing the movements of empty freight cars with the objective of minimizing cost, while satisfying the demand and supply for empty cars. Moreover, because there are many types of freight cars and not every one of them may be used to transport all types of goods, empty cars may have to be delivered to a terminal even if there is already an excess of empty cars at that terminal. Empty transports thus represent a substantial part of the total workload in a railway system, and an efficient empty car distribution process is very important for reducing both operating and capital costs.

Some of the costs for empty transports in a railway system are proportional to the number of freight cars moved and directly related to transportation distances. Other costs, such as the classification costs, are caused by the freight car handling at yards, and constitute a large portion of the total cost for empty distribution, especially in smaller systems. The costs for car-handling operations often depend on the number of *car groups* that are handled at the yards, where a car group corresponds to one or more cars that are held together during the operations and therefore can be treated as one handling unit. If fewer but larger groups of cars could be handled at yards, not only would the shunting and classification processes be simplified, but the associated per-unit car cost would also decrease. Thus, the cost structure displays an economy-of-scale behavior. However, most models for empty freight car distribution, such as those reviewed in Cordeau et al. (1998) and Dejax and Crainic (1987), do not explicitly take into account these economies of scale. As a result, these models may yield freight car distributions that involve movements of many small car groups, and larger operating costs.

The purpose of this paper is to formulate and solve an optimization model for empty freight car distribution that explicitly considers the economies of scale in the distribution process. The railway system we consider is a scheduled railway in which all trains are running according to a timetable. The timetable defines arrival and departure times, capacities of trains, which days of the week the trains are run, etc. We assume that there exists information about available transportation capacities in the trains and also reservation possibilities for freight cars in the trains. The timetable, together with the reservation possibilities, certifies that the empty transports can be performed as planned, making the implementation phase of the empty car distribution deterministic. The modeling of the economies of scale is performed by defining fixed costs for some arcs in the network, in addition to the arc costs proportional to flow. The proposed model is dynamic (time dependent) in the sense that a time dimension is included in the model, with multiple time periods considered simultaneously.

The model will generate distribution plans with coordinated car movements so that the number of car groups and the number of car-handling operations are reduced. Thus, the total distribution cost will decrease. The optimization model is intended for operational planning, but it can also be used at the tactical and strategic planning levels. The motivation for the work comes from Green Cargo, formerly the freight transportation division of Swedish Railways (SJ), and the planning situation at this company is the background to the formulation of the proposed model. A number of assumptions about the problem require that the information systems at the railway company be rather well developed, which is the case at Green Cargo. Although we have the Swedish situation in mind, we are convinced that the problem scenario is similar in many other railway companies, and that the modeling approach is applicable to general settings, especially for railways that run trains according to detailed prespecified schedules.

The model gives rise to a capacitated, multicommodity network flow optimization formulation with fixed costs associated with paths defined on the multiperiod time-space diagram representing the planning problem. It belongs to the class of network design formulations that is known to be difficult to address (Magnanti and Wong 1986, Minoux 1986, Balakrishnan et al. 1997). Indeed, our computational results show that even for the moderate dimensions of the Green Cargo, the branch-and-bound algorithm of a state-of-the-art commercial software cannot identify feasible solutions within the time constraints inherent to the planning process. We propose, therefore, a tabu search metaheuristic that takes advantage of the structure of the problem and that yields good-quality solutions in reasonable computing times.

To summarize, the contributions of the paper are: the analysis of economies of scale in empty freight car distribution, the time-dependent network optimization model for the empty freight car distribution problem in scheduled railways that explicitly takes into account these economies of scale, and the tabu search metaheuristic that efficiently identifies good solutions to the model.

The remainder of the paper is organized as follows. Section 1 describes the empty freight car distribution problem and presents a brief literature survey. The cost structure of the process is discussed in §2. Section 3 is dedicated to the description of how we model the economies of scale and to the presentation of the full mathematical formulation of the corresponding model. Section 4 describes the tabu search heuristic. Numerical results are presented and analyzed in §5. Finally, the paper is concluded in §6.

1. Problem Description

In this section we describe the empty freight car distribution problem in a scheduled railway. The general planning process for empty freight car distribution is described, and the impact of economies of scale on planning is discussed. Related literature is reviewed.

1.1. The Planning Process

Planning activities at a railway company may be divided into three planning levels: Strategic, where the size of the resources is determined; tactical, where the allocation of resources is planned; and operational, which schedules, monitors, and manages the actual operations. In this paper, we focus on issues at the operational planning level.

The flow distribution of loaded and empty cars in a railway network constitutes a *transportation pattern*. It includes the specification of the origin and destination of all freight car movements, the assignment of freight cars to trains, and the temporal aspects of the transport. The operational costs strongly depend on the transportation patterns that, given the available system resources, result from the tactical planning activities. Thus, one has to study the correlation between the tactical and operational planning levels.

The most important part of tactical planning in a scheduled railway is to construct the timetable. The timetable defines days of operation, routes, stops, arrival and departure times, and capacities of the trains. It also defines which blocks are built at yards and which train connections are possible. The handling times at classification yards are taken into account in the timetable. As a consequence, the timetable defines all feasible ways to perform the transport of cars from an origin to a destination. The transportation time on each route between two terminals, including handling times at terminals and classification yards, is also known. The timetable is usually revised a few times a year.

We define a train by its route, origin, destination, intermediate stops and physical path, as well as the day of operation, and the departure and arrival times at all stations. This means that if the timetable specifies that one train shall run a specific route each day, we consider this as multiple trains, one for each day. Thus, each train is run only once.

The general freight transportation pattern is determined simultaneously with the timetable during tactical planning. The train capacities are determined at the tactical planning level, and this capacity has to be shared between loaded and empty freight cars. One of the most important decisions in operations planning, then, is to update, instantiate, and implement the transportation pattern, according to the timetable, for the actual demand.

As soon as the demand for loaded transport becomes known, the corresponding loaded freight cars are booked on the trains. Most of the demand is known at least one day before the loaded transport is to be performed. The residual train capacity after loaded cars are booked can be used for empty cars. The loaded transports cannot be altered to accommodate empty distribution, so the train capacities for empty freight car transport are fixed.

The loaded transports generate a demand for (and a future supply of) empty freight cars. The need for empty freight cars is known at least one day before demand for cars has to be fulfilled, since the loaded transport demands are known one day in advance. In smaller railway systems such as Green Cargo, the span between the instant at which a demand for empty cars is known and the time when it has to be fulfilled is longer than most transportation times, which means that the demand for empty cars can be considered as known. The objective of the operational planning of the empty freight car distribution is to satisfy as many demands for empty cars as possible, while minimizing the distribution costs.

Compared to loaded transports, the planning of empty transports gives more freedom to the railway company in how to implement the empty movements. A transportation order of loaded cars has a fixed origin and a fixed destination and is often scheduled for a specific day, which means that there are very few alternatives for implementing the order and the movement. When repositioning empty cars, there are no origin-destination specific demands that have to be fulfilled, and no specific time schedule for each empty transport. The only requirement is to satisfy the total demand, given the supply for each car type at each terminal. The freedom in determining how to move empty freight cars can be used to coordinate the empty transports in order to obtain a cost-efficient transportation pattern that explicitly accounts for economies of scale.

The coordination idea is illustrated in Figure 1, using a small railway network with only four terminals (A, B, C, D) and two classification yards (E, F). At Terminals A and B there is a supply of empty cars of four and one units, respectively, and at each of Terminals C and D there is a demand for empty cars of two units. The distances between the terminals and yards are proportional to the lengths of the arcs in the figure. The total transportation distance will be minimized if two cars are transported from A to C, one car from A to D, and one car from B to D. An alternative solution is to send two cars from A to C and two cars from A to D. The total distance transported will then increase, but this transportation pattern will simplify the work (car handling) at terminals B, E, and F. If the increase in the transportation distance is not too large, many railway practitioners would prefer the second solution. In this paper we develop an optimization model for empty freight car distribution that reflects the relative gains associated to this type of consolidation/coordination of empty freight car movements.

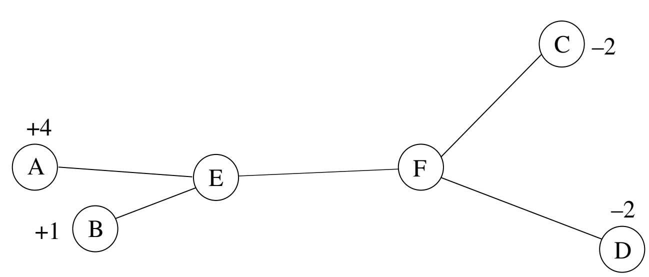

Figure 1 Illustration of Transportation Trade-Offs

1.2. Literature Review

Several models for empty freight car distribution have appeared in the literature. See Cordeau et al. (1998) and Dejax and Crainic (1987) for reviews.

Most of the models for empty distribution do not explicitly take into account economies of scale, however. Instead, it is implicitly assumed that transports are organized at the tactical planning level to utilize some large-scale advantages. This means that the transports are scheduled so that each origin-destination specific movement is implemented in a cost-efficient way. The routing of the freight cars on trains is not explicitly included in most models. For example, Powell and Carvalho (1998b) determine the number of flat cars to be sent from one terminal to another by formulating a logistics queuing network (Powell et al. 1995, Powell and Carvalho 1998a). Various aspects are considered in the decision process, but it is assumed that the cost of sending cars is proportional to the number of cars. Holmberg et al. (1998) present a more detailed description of the transportation possibilities in a railway system. They include the assignment of cars to trains in the model, but they do not consider the large-scale impacts of consolidation/coordination in car distribution.

Haghani (1989) formulates a model for the simultaneous tactical planning of train routing, train makeup, and empty freight car distribution. The estimated flow of empty cars is considered in the tactical planning and influences the train routing and makeup decisions. In this model, the economies of scale and empty distribution are taken into account at the tactical level.

Turnquist (1994) explicitly takes economies of scale in empty freight car distribution into consideration. Turnquist formulates the empty freight car distribution problem as an uncapacitated network design problem. The design arcs correspond to the origin-destination (OD) pairs utilized in the distribution. The objective is to find a distribution that combines minimizing the transportation distance and minimizing the number of OD pairs for which empty cars are sent.

Comparing our objectives and model to these works, it appears that while Haghani (1989) approaches the problem from the tactical planning level, we have the operational planning stage in mind, taking the trains and their schedules as stipulated. Furthermore, while Turnquist (1994) does not include any transportation capacities and models each possible OD flow with an arc, we explicitly consider train capacities. We also acknowledge that several trains may be used to perform a transport between an origin and a destination and, conversely, that cars moving between many OD pairs may be transported in the same train. This makes it impossible to model train capacity with simple flow restriction on an arc representing an OD flow. Train capacities are a central planning issue at many railway companies (see Holmberg et al. 1998 for example) and the model we propose contributes to address it in the context of empty car distribution.

2. The Cost Structure

In order to design an efficient transportation pattern, it is necessary to analyze the cost structure of the empty freight car distribution process and to identify the relevant costs that directly or indirectly are associated with a given pattern. We are interested in costs that can be influenced by changing the transportation pattern, i.e., costs that depend on the flow and routes of freight cars. Costs that do not depend on the transportation pattern design (e.g., maintenance costs, some personnel costs, capital costs) are not considered. Moreover, to simplify the presentation, we assume holding costs to be independent of the terminals where cars are held (this is generally true, since holding costs are usually dependent on the car inventory cost and not on the particular location). We may thus avoid including them in the formulation.

In this section, we describe and analyze the most important costs for the empty freight car distribution. We first describe the costs corresponding to trains, and then the costs for moving cars that are directly related to transportation distance. The latter is included in most models for empty freight car distribution. We then describe the car-handling costs related to classification at yards. Throughout the section, we distinguish between costs that depend on decisions at the tactical planning level from those that depend on the operational planning.

2.1. Train Costs

An important share of the total costs for implementing a transportation pattern is associated with the trains. Some train costs are fixed and do not depend on what is transported in the train, for example, engine and driver costs. Such costs are related to the decision to run a train or not. Other costs, such as energy costs, depend on the load of the train and the transportation pattern. We denote these variable train costs as transportation costs, and discuss them below.

Train costs can be modeled as fixed costs, positive if the train is run, zero otherwise. In a scheduled railway, most trains are determined at the tactical planning stage, and only a very minor part of the total train cost can be influenced by the transportation pattern that is determined at the operational planning stage. Since the empty car optimization model is aimed at operational planning, we ignore these fixed train costs in the model, assuming that they are determined at the tactical planning stage.

2.2. Transportation Costs

A main component of total cost, transportation costs depend on the empty freight car mileage and include, for example, energy costs and freight car wear costs. Transportation costs can be modeled as proportional to the number of moved cars and the transportation distances. Tactical planning impacts transportation costs through the design of the timetable, while the operational planning stage does so by changing the routes of the freight cars.

Transportation costs are of particular interest to the distribution of empty cars, since the origin-destination flows of empty cars are not given, but can be influenced during operational planning. We thus include a transportation cost that is proportional to the number of cars moved for each origin-destination pair, and the corresponding distance.

2.3. Yard Costs

It is important to distinguish the costs associated with classification yards according to the planning level at which they are generated. Strategic planning determines the existence of classification yards, which is certainly associated with a large cost. Tactical planning determines which blocks are built at each yard, while operational planning specifies how the blocks are built. Consequently, at the operational planning level the only costs of interest are the costs that depend on the work—the freight car handling—that has to be done in the yard. We do not include any strategic or tactical yard costs in the optimization model, but we do include a cost representing the operational yard costs associated to freight car handling, as described hereafter.

Yard operations are very complex, and generally give rise to equally complex costing formulae. For the purposes of the empty car distribution problem considered in this paper, one may simply consider that the objective of yard operations is to sort and group cars to form the outgoing trains. To represent the associated costs, we define the notion of *car cluster*, which is closely related to the more common one of *block*.

- A *block* is a group of cars that moves together between two yards. A block moves on trains and through yards without any manipulation on the cars that make it up. The origin yard of the block does not have to be the origin of all cars in the block, and the destination of the block does not have to be the final destination of all cars in the block. Thus, cars travel in a block during *part* of their transport. Which blocks can (or have to) be built at a yard depends on the design of the yard at the strategic planning level and on the design of the timetable at the tactical level.
- A *car cluster* is a group of cars that has a common initial yard and a common final destination yard and that are moving together on trains and through yards during the *whole* transport. Cars in a car cluster are physically connected all the way from the origin to the destination, and they never have to be separated during the transport. Thus, one car cluster can be considered as one handling unit in yard operations, causing one unit of cost. The number of car clusters is determined at the operational planning level and is given by the transportation pattern. Since trains are defined to be time-and-day-specific (each train exists only once), each car cluster exists only once. The cars in the car cluster can be loaded or empty.

When a train arrives at a yard, the composing cars are classified with respect to outgoing trains. The incoming trains often have to be split into many small car groups in order to build the outgoing trains. In the worst case, each car group consists of one car cluster. The operational costs at a yard can then be described as consisting of two components:

- A car-handling cost proportional to the number of cars handled at the yard.
- A group-handling cost proportional to the number of car groups that are handled at the yard.

In the model presented in this paper, yard costs are modeled as follows. The car-handling cost is determined by the route of the car. The sum of the car-handling costs related to all cars on a given route is included in the transportation cost (that is thus proportional to the number of cars moved on that route). For the group-handling cost, we approximate the number of car groups handled at a yard by the number of car clusters that pass through the yard. It is an approximation because when two car clusters use the same sequence of trains for part of their routes, the car clusters can be consolidated for their common journey and avoid being disconnected at intermediate yards. In this case, the number of car-handling operations will not be strictly equal to the number of car clusters. Differences are not enough, however, to prevent the number of car clusters from being a good basis for the evaluation of the yard effort and cost (Joborn 2001). Consequently, a (fixed) cost is associated with each car cluster, equal to the sum of the group-handling costs of all the yards through which the car cluster passes.

In railway systems that are not strictly scheduled, handling times at classification yards are of major concern. In these cases, it should be even more interesting to find transportation patterns that reduce the yard workloads. Handling times often depend nonlinearly on workloads, however. Therefore, for such railway systems, another cost model should be used. For the strictly scheduled railways we consider in this paper, the cost model we propose is, nevertheless, appropriate.

3. Mathematical Model

In a strictly scheduled railway system, the timetable is determined at the tactical planning stage. Consequently, arrival and departure times of trains are fixed, available train capacities are known, and the possible routes in the railway network are stipulated. Transportation times and handling times at classification yards are given by the design of the timetable. The timetable does not have to be cyclic, which means that the possible movements may differ from period to period.

The loaded transports, either known or given by a forecast, are the input of the empty distribution planning process. As a consequence, the supply and demand for each car type at each terminal are known. Each demand can only be fulfilled by one car type. We assume that there always exists a way to satisfy all demands and that demand cannot be backordered (the demand that is possible to fulfill can be precalculated as described, for example, by Holmberg et al. 1998). The demand for loaded transports is known (and booked) before the empty distribution is planned, and loaded transports are considered to be unalterable. Thus, a deterministic approach may be used in deciding how to distribute the flow of the empty freight cars. The supply, demand, transportation possibilities (trains), and transportation capacities may vary from day to day, and thus a time-dependent problem representation is required.

In the following, we present the fundamental modeling ideas in representing the time-dependent elements of the problem and constructing the space-time network. We then introduce the mathematical model that aims to determine origin-destination-specific movements of empty freight cars, together with their routes through the railway system, at minimum cost over a given planning horizon.

3.1. Model Specification and Assumptions

The model formulation is based on a time-expanded network, called the *train network*, that describes the set of *train* movements, denoted G, during the planning horizon (see Figure 2). The planning horizon is divided into T time periods (days). Each terminal $n \in N$ is represented by at least one node in each time period, with additional nodes being inserted to account for trains, if needed. There are two types of arcs in the train network: train arcs that correspond to elements of G and *inventory* (or *holding*) arcs that connect the nodes corresponding to copies of the same terminal in successive time periods.

The flow in this network represents *empty* freight cars moving from one terminal to another or being held in inventory. Since we assume that there are several different types of nonsubstitutable car types, we must consider multicommodity flows. We denote by K the set of car types (commodities). Note that each commodity may have several sources and sinks and that a node can be both a source and a sink for commodities. To simplify the presentation, we assume that all car types have the same size (length and weight) and, thus, the train (residual) capacities can be expressed in numbers of empty freight cars. Consequently, an arc representing a train $g \in G$ has a *bundle* capacity u_g on the total number of cars, of all types, that can be moved by that train. Supplies and demands of empty cars are assumed to be realized at the beginning of each time period. The net supply (demand, if negative) of cars of type $k \in K$ at terminal n at the beginning of period $t = 1, 2, \ldots, T+1$ is denoted b_{nt}^k.

As defined in §2.3, a car cluster is a set of cars that are held together all the way from their initial origin to their final destination, which means that the cars in a car cluster are transported on the same day in the same trains. This means that each car cluster q can be characterized by a unique set of trains H_q that are used for its transport. The trains in H_q form a *path* in the train network. To illustrate these notions, consider the two train paths depicted with dashed lines in Figure 2. One path originates from Node 1, representing Terminal A in Time Period 1 and terminates at Node 5, representing Terminal C in Time Period 4. The other path originates from Node 2, representing Terminal A in Time Period 2, and terminates at Node

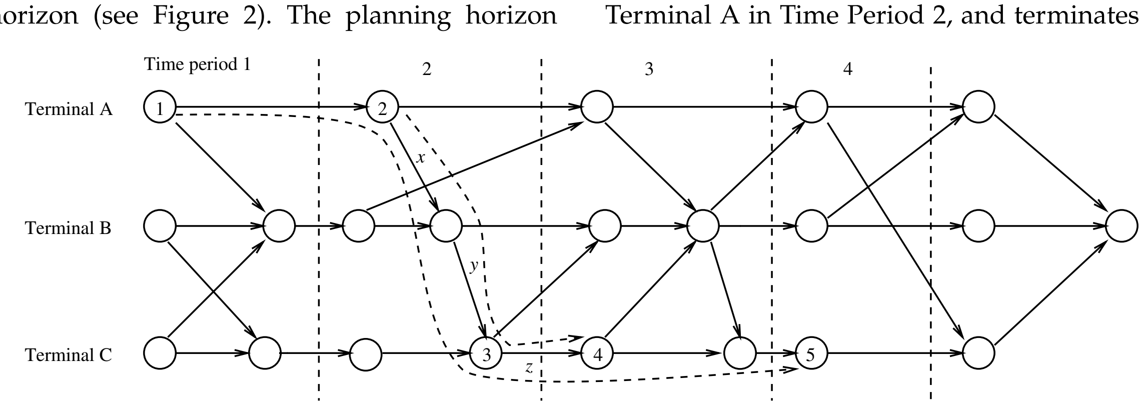

Figure 2 The Train Network and Two Paths

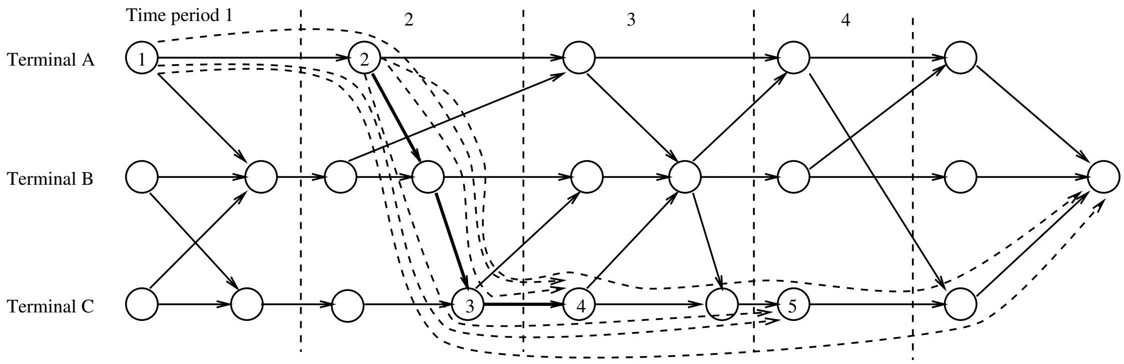

Figure 3 The Train Network: One Kernel Path and All Corresponding Paths

4, representing Terminal C in Time Period 3. Both these paths include the arcs x and y, representing trains from A to B and from B to C, respectively. The flows on both these two paths represent the transport of empty cars from A to C by the same trains (at the same period). Thus, the origin and destination terminals are the same, as well as the spatial movement, the only difference being when the cars become available (different temporal source nodes) and when they are required to fulfill demand (different temporal sink nodes). These cars thus belong to the same car cluster.

More than two paths in the train network include a given set of trains H_q corresponding to car cluster q. In fact, all paths that start at the origin of the first train arc at or before its departure time and that end at the destination of the last train arc after its arrival time belong to this set, denoted P_q. Figure 3 displays all such paths for the train set and car cluster of Figure 2. These paths have three arcs in common: the train arcs (x, y in Figure 2) and one inventory arc (arc z). Note that demand is realized in the beginning of the time periods and, consequently, Node 3 cannot be a sink node and no path can terminate at Node 3.

We therefore define, for each subset of paths P_q that corresponds to a unique set of trains (and train arcs) H_q, a *kernel path* p_q as a path consisting of all the arcs (train arcs and inventory arcs) that are common to all paths in the set (the chain of bold arcs in Figure 3 or, equivalently, the chain x, y, and z in Figure 2). A kernel path p_q differs from the paths in its corresponding set P_q by the fact that it does not include inventory arcs before the departure of the first train in H_q, nor any inventory arcs after the beginning of the time period following the arrival of the last train in H_q. Each kernel path p_q thus has a unique origin terminal, destination terminal, departure time, arrival time, and corresponds to a unique set of trains, H_q.

Each kernel path represents one way of performing a transport between two terminals, and therefore it represents exactly one car cluster. All kernel paths emanate and terminate at the beginning of the time periods because the supply and demand are assumed to be realized at the beginning of the time periods. The flow on a kernel path corresponds to the total number of empty freight cars, of all car types (commodities), to be transported in one car cluster between two terminals in a specific combination of trains. Consequently, the flow on a kernel path can originate from different origin nodes representing the same supply terminal in different time periods, and can terminate at different sink nodes, representing the same demand terminal in different time periods.

A kernel path has no capacity restriction per se. There are, however, the capacities u_g on the train arcs that are included in the kernel path. The train capacity restriction is thus a bundle constraint over the sum of the flows of all kernel paths that include the corresponding train arc. The cost for sending a car cluster is equivalent to a fixed cost associated with the corresponding kernel path. The flows on the paths in P_q thus generate only one fixed cost because they correspond to exactly one car cluster q. Figure 4 displays all possible kernel paths of the train network in Figure 2. The possible kernel paths are given by the transportation structure of the railway system. In the railway systems we study in this paper, we consider the kernel paths as generated beforehand, defined by the timetable.

Given the kernel paths, the train network is transformed into another time-dependent structure, denoted the *kernel network* and illustrated in Figure 5. Each kernel path in the train network yields an arc in the kernel network. In the latter, there is only one node for each terminal and time period, and the possible train connections are implicitly defined by the kernel arcs. To each kernel arc, one associates a variable flow cost representing the per-car transportation cost, as well as a fixed cost that is incurred if a nonzero flow is assigned to the arc. Bundle capacity restrictions limit the total flow on all kernel arcs that include the same train. Moreover, a kernel arc may be included in several capacity constraints, since it may correspond to movements on several trains.

The empty flow distribution problem thus appears as a particular variant of the capacitated, fixed

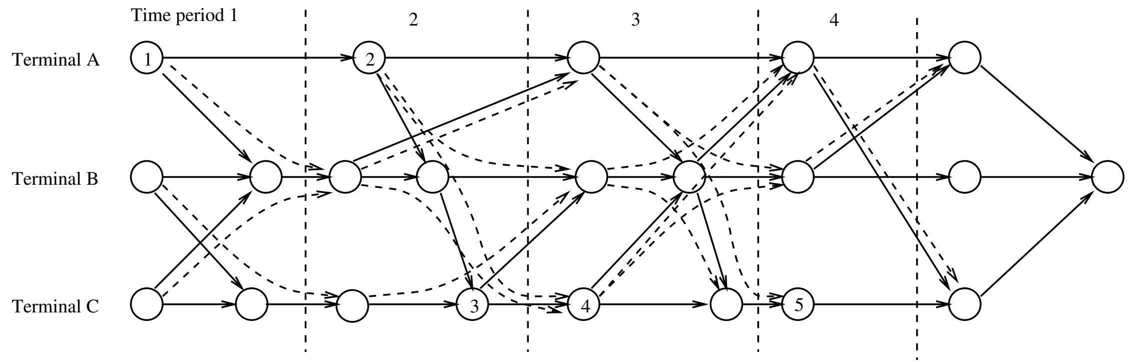

Figure 4 **Train Network with All Kernel Paths**

cost, multicommodity network design problem class (Magnanti and Wong 1986, Minoux 1986, Balakrishnan et al. 1997, Crainic 2000), where one has to determine which kernel arcs to include in the solution. The major difference between our formulation and the classical network design models is that, in the latter, capacity constraints are defined for the total flow on design arcs, whereas in our model, each capacity constraint represents a limit on the total flow on *several* design arcs. This adds to the complexity of what is known to be an NP-Hard problem. In the following, we present the model formulation that indicates precisely the problem one must solve.

3.2. Model Formulation

Let P represent the set of all kernel arcs in the network and define:

P^o_{nt}, P^d_{nt}: Set of kernel arcs respectively emanating from and terminating at terminal n in period t, P^o_{nt}, $P^d_{nt} \subset P$, $n \in N$, $t = 1, 2, \ldots, T+1$, ($P^o_{n,T+1} = P^d_{n1} = \varnothing$).

P_g: Set of kernel arcs that include train g, $g \in G$, $P_g \subset P$.

m_p: Maximum number of empty cars that can be sent in the car cluster corresponding to kernel arc p, $p \in P$.

f_p: Fixed cost for sending a car cluster by using the trains corresponding to kernel arc p, $p \in P$.

c_p: Variable (unit) cost for sending a car on the trains corresponding to kernel arc p, $p \in P$.

v^k_n: Salvage value of a car of type k at terminal n at the end of the time horizon, $n \in N$, $k \in K$.

Given initial supplies and demands for each commodity k, at each terminal n, denoted i^k_{n0}, we define three sets of decision variables:

x^k_p: Number of empty cars of type k transported on the trains corresponding to kernel arc p, $p \in P$, $k \in K$.

$$y_p = \begin{cases} 1 & \text{if any cars are sent in the car cluster} \\ & \quad \text{corresponding to kernel arc } p, p \in P; \\ 0 & \text{otherwise.} \end{cases}$$

i^k_{nt}: Number of cars of type k in inventory at terminal n at the end of period t, $n \in N$, $k \in K$, $t = 1, 2, \ldots, T+1$.

The model may then be written as

$$F(y, x) = \min \sum_{p \in P} \left(f_p y_p + \sum_{k \in K} c_p x^k_p \right) - \sum_{n \in N} \sum_{k \in K} v^k_n i^k_{n,\,T+1}, \tag{1}$$

subject to

$$i^k_{nt} = i^k_{n,\,t-1} + b^k_{nt} + \sum_{p \in P^d_{nt}} x^k_p - \sum_{p \in P^o_{nt}} x^k_p \quad \forall k \in K,\ \forall n \in N,\ t = 1, 2, \ldots, T+1, \tag{2}$$

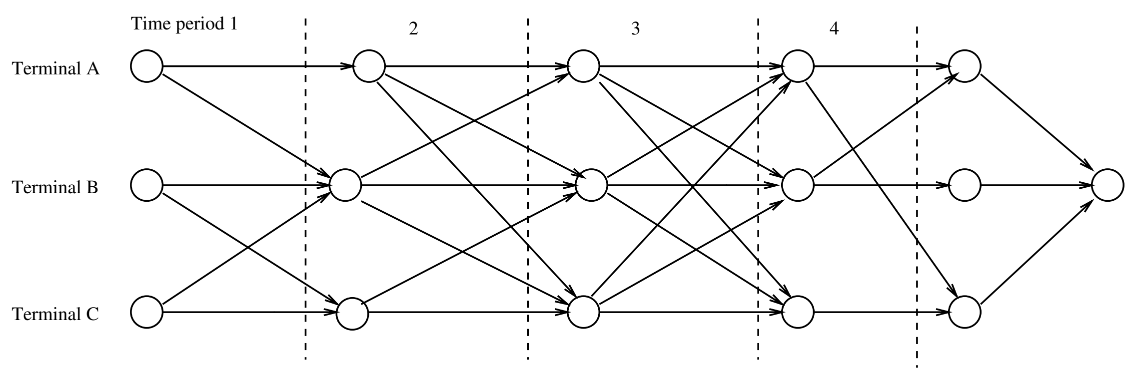

Figure 5 **Kernel Network Corresponding to Kernel Paths in Figure 4**

$$\sum_{k\in K} x_p^k \le m_p y_p, \quad \forall p \in P, \tag{3}$$

$$\sum_{p\in P_g}\sum_{k\in K} x_p^k \le u_g, \quad \forall g \in G, \tag{4}$$

$$x_p^k \ge 0, \quad \forall k \in K, \ \forall p \in P, \tag{5}$$

$$y_p \in \{0, 1\}, \quad \forall p \in P, \tag{6}$$

$$i_{nt}^k \ge 0 \quad \forall k \in K, \ \forall n \in N, \ t = 1, 2, \ldots, T+1. \tag{7}$$

The objective is to minimize the total costs of the distribution process. The objective function (1) has two parts. The first part represents the car cluster costs for utilizing kernel arcs and the corresponding variable transportation costs. The second part captures the costs associated with the "last" period of time horizon and the utility of having cars at different terminals (represented by the salvage values). See Hughes and Powell (1988) for a thorough analysis of end-of-horizon effects in dynamic vehicle-allocation models.

Constraints (2) represent inventory balances, stating that the number of cars in inventory of each car type at each terminal is equal to the outgoing inventory of the previous time period plus supply of cars, plus cars arriving from other terminals, minus cars sent to other terminals. Relations (3) are linking constraints and state that the binary variable y_p has to be one if flow is assigned to the kernel arc p (i.e., a car cluster exists if it is used to move flow). Each constraint of Type (4) bounds the total flow of all commodities on all kernel arcs by the capacity of the associated train arc. It is the set of Constraints (4) that makes our model different from, and more difficult than, classical network design formulations.

4. Solution Method

Network design problems are known to be difficult, and our variant is no exception, as emphasized by the computational results presented in the next section. We therefore developed a tabu search heuristic (Glover and Laguna 1997) to address the empty freight car distribution model described in §3. In this section, we describe the main ideas and elements of the method. More details may be found in Crainic et al. (1997). The tabu search procedure is summarized in Table 1.

A solution in the search space of the tabu search procedure is defined as a binary vector y that indicates whether each kernel arc is open or closed. A solution is evaluated by solving the subproblem $P(\bar{y})$ that results from fixing all binary variables in Model (1)–(7) to $\bar{y}$. $P(\bar{y})$ is a linear multicommodity network flow problem that is solved by an LP-solver. The objective function value of $P(\bar{y})$, denoted $z(\bar{y})$, includes transportation costs, but does not include the design costs of the kernel arcs (car cluster costs). The value $F(\bar{y})$ of solution $\bar{y}$ is then obtained as $z(\bar{y})$ plus the costs of the binary variables $\bar{y}$ that are fixed to one.

We define two types of neighbourhoods. For a solution $\bar{y}$, neighbourhood $N_1(\bar{y})$ has one additional kernel arc closed, while neighbourhood $N_2(\bar{y})$ has one additional kernel arc open. Hence, a move corresponds to changing the value of exactly one binary variable y_p by either opening or closing the associated kernel arc p. Neighbours are evaluated by solving the multicommodity flow subproblems with the appropriate changes in the binary variables. We move to a neighbour y' of type $N_2(\bar{y})$ only if $z(y')$ is strictly better than $z(\bar{y})$. This follows from the fact that opening a kernel arc should improve the flow distribution. If this is not the case, there is no incentive to move to y'.

Generally, the number of kernel arcs is very large, while only a small number have nonzero flows. Hence, the neighbourhood $N_1(\bar{y})$ is small, while neighbourhood $N_2(\bar{y})$ is very large. It is therefore intractable to calculate the exact value of each neighbour in $N_2(\bar{y})$. Instead, we first calculate a *target* (estimated value) for the change in the objective func-

Table 1 The Tabu Search Procedure

Parameters

- v_1, v_2: Number of neighbours to examine in $N_1(\bar{y})$ and $N_2(\bar{y})$, respectively.
- l_{open}, l_{close}: Tabu tags (in number of iterations) forbidding the reversal of the decision on an opened and closed arc, respectively.
- i: Number of iterations between artificial arc cost updates. α, β: Growth and decrease factors for artificial arc costs, respectively.

Step 0. Select a feasible starting solution $\bar{y}$. Set the currently best solution $y^* := \bar{y}$. Initialize iteration counter $t := 1$.
Step 1. Randomly select v_1 neighbours in $N_1(\bar{y})$. Calculate the corresponding objective function values F of the neighbours.
Step 2. Calculate the target value of all neighbours in $N_2(\bar{y})$. Calculate the objective function value F of the v_2 neighbours with the best target values.
Step 3. Sort all the evaluated neighbours y into a list L according to $F(y)$, best neighbour first.
Step 4. Pick the first neighbour y' in L.
Step 5. If y' is tabu, and either it corresponds to an infeasible solution or $F(y') \ge F(y^*)$, remove y' from L and go to Step 4.
Step 6. If y' is a neighbour of type $N_2(\bar{y})$ and $z(y') \ge z(\bar{y})$, remove y' from L and go to Step 4.
Step 7. If y' is a feasible solution and $F(y') < F(y^*)$, set $y^* := y'$.
Step 8. Update tabu lists according to the opened or closed kernel arc in y', and by (potentially) making some kernel arc nontabu.
Step 9. If t is a multiple of i, update artificial arc costs.
Step 10. Move to the neighbour y', set $\bar{y} := y'$, increase iteration counter $t := t + 1$, go to Step 1.

tion value when moving to a neighbour in $N_2(\bar{y})$. The neighbours with the best target values are then evaluated exactly.

The target value $\bar{f}_p$ for opening a kernel arc p, emanating from node i and terminating at node j, may be calculated in many ways. We experimented with several approaches, all of them based on the dual variables obtained from the solution of the linear subproblem $P(\bar{y})$:

(1) Reduced cost of the y_p variables in the LP-relaxation of model: $\bar{f}_p = f_p + m_p(\min\{0, \bar{c}_p - \sum_{g\in P_g} \delta_g)$, where δ is the dual variable of Constraint (4) and $\bar{c}_p = c_p - \max_k\{d_j^k - d_i^k\}$, with d_i^k standing for the price of node i for commodity k (the dual variable of the associated constraint (2)). Note that m_p can be considered as an (over-) estimate of the flow on the kernel arc if it is opened.

(2) Modifying the target value of Type 1 by replacing m_p with an estimate of the flow of each commodity if the kernel arc is opened, computed as the minimum of inflow of node i and outflow of node j.

(3) Using $\bar{f}_p = f_p + \bar{c}_p$ as target value. This corresponds to an estimate of the change in the objective function value if one unit of flow is assigned to the opened kernel path. In the calibration tests, this approach gave the best results and is used in this paper.

To speed up the tabu search, the neighbourhood $N_1(\bar{y})$ is not entirely explored. Rather, a random sample of the neighbours in $N_1(\bar{y})$ is evaluated. The move is then made to the best-evaluated neighbour of type $N_1(\bar{y})$ or $N_2(\bar{y})$.

There are two different short-time tabu lists. One contains the kernel arcs that were recently opened; the tabu restriction is to not close them. The other list contains the kernel arcs that were recently closed, and it is tabu to open them. The evaluation function is used as an aspiration criterion and the tabu restrictions are overridden if the move leads to a new best solution.

Artificial arcs are added to the network, such that the problem always has a solution irrespective of the set of open kernel arcs. The artificial arcs are added from the super sink beyond the time horizon to the first node at the beginning of the planning period of each terminal. In this way, we can "tunnel," i.e., we can move from one feasible solution (feasible with respect to the original problem) to another feasible solution via a set of infeasible solutions. The costs of the artificial arcs are regularly lowered and raised during the search in order to enable tunneling and provide feasible solutions, respectively. Two strategies for updating the cost of the artificial arc costs have been analyzed. Both update artificial arcs costs every ith iteration.

(1) All artificial arc costs are multiplied (divided) by two if the solution has been infeasible (feasible) for all the i last iterations. This strategy was proposed by Gendreau et al. (1994).

(2) The second strategy modifies Strategy 1 by multiplying the cost on one individual artificial arc with a growth factor α ($\alpha > 1$) if there has been any flow on that arc in any of the i last iterations. The costs on all artificial arcs are multiplied by a decrease factor β ($\beta < 1$) if the solution in all the last i iterations has been feasible.

In the calibration tests, Strategy 1 resulted in frequent but small improvements, while Strategy 2 resulted in less frequent but larger updates of the current best solutions. This is a result of the finer tuning of the artificial arc costs in Strategy 2, which allows many ways of tunneling, making it slower to find a feasible solution, but resulting in feasible solutions of higher quality. After one hour of computation time in the tests, the solutions provided by Strategy 2 were often better than those of Strategy 1, and Strategy 2 is used in this paper.

An initial solution is found by first solving the problem with all fixed costs set to zero; i.e., the variable transportation costs are minimized. All kernel arcs that have a nonzero flow are then considered open. This solution is improved by closing the open kernel arcs one by one. If the optimal variable cost is not changed by the closure of one kernel arc, this kernel arc is kept closed; otherwise, the kernel arc is reopened.

5. Computational Results

The objective of the numerical experiments is to analyze the performance of the proposed solution method and to verify that problems of real-life dimensions are solvable in reasonable time. The experimental results also show that the distribution plans obtained by using the proposed model indeed use economies of scale.

For confidentiality reasons, data from Green Cargo were not directly available. Instead, we randomly generated problem instances that display structures closely similar to those existing at Green Cargo. Thus, we assume that trains run in a hub-and-spoke structure, corresponding to major classification yards and terminals. All terminals are assigned to exactly one major classification yard, and there are one to three daily trains between a classification yard and its terminals, as well as between the major classification yards. Train capacities are randomly generated around a day-specific mean value chosen to ensure that transportation capacities are reasonably scarce. For each freight car type, each terminal is first assigned to be a demand or a supply terminal of different levels, and then supplies and demands

Table 2 Dimensions of Test Problems

Prob	Hubs	Terminals	Commodities	Time Periods	Kernel Arcs 0/1 Variables	Continuous Variables	Constraints
*p*1	3	6	2	3	94	342	286
*p*2	3	6	2	3	118	414	334
*p*3	3	6	2	3	262	654	643
*p*4	4	12	5	4	710	4620	4027
*p*5	4	12	5	4	728	4360	4117
*p*6	4	12	5	4	849	5454	4722
*p*7	6	6	10	5	396	4776	4540
*p*8	6	6	10	5	401	4831	4590
*p*9	6	6	10	5	1288	14020	13603
*p*10	6	25	10	5	4092	46762	43020
*p*11	6	25	10	5	6331	71391	65410
*p*12	6	25	10	5	12061	124670	122867
*p*13	15	25	15	5	21701	349841	329305
*p*14	15	25	15	5	34036	547201	514330
*p*15	15	25	15	5	47923	769393	722635
*p*16	5	50	20	5	15799	338779	323550
*p*17	5	50	20	5	18084	386764	369250
*p*18	5	50	20	5	18541	396361	378389
*p*19	25	25	15	4	1351337	21623642	20274720
*p*20	25	25	15	4	1996638	31948458	29954235
*p*21	25	25	15	4	2079694	33277354	31200075

are randomly generated according to these base levels. The salvage value for empty freight cars at the end of the planning period is set to zero in all tests presented in this paper. Full description of the generation procedure may be found in Joborn (2001).

Twenty-one (21) problem instances were generated. Table 2 displays the problem dimensions, as well as the number of variables and constraints in the corresponding formulation. Three problem instances are tested for each specific combination of number of hubs, terminals, commodities, and time periods in the planning period. These instances differ in their geographical layout, number of trains and their capacities, and supply and demand structure are varied. Test problems *p*1 to *p*15 are increasingly larger and allow comparisons to a state-of-the-art optimization solver. Problems *p*7 to *p*9 and *p*19 to *p*21 are networks without the hub-and-spoke structure ("only hubs and no spokes") and allow observation of the performance of the proposed methodology on cases with different network layout and dimensions. Problems *p*16, *p*17, and *p*18 correspond to the size of the Green Cargo network and, as indicated previously, display the basic structure and operation characteristics of the company. Thus, these problems stand for the "real-life" instances.

The tabu search algorithm is implemented in the programming language C and CPLEX 6.5 (ILOG 1999) is used as LP-solver in the search algorithm. Tests are performed on a multiuser SUN Ultrasparc 2/2200 with 2 Gb RAM. To evaluate the performance of the tabu search algorithm, we also attempted to solve the test problems with the mixed integer programming (MIP) solver CPLEX (version 6.5).

Experiments with different parameter settings for the tabu search procedure are presented in Crainic et al. (1997). Table 3 summarize the evaluated and the best values (column "Best") for the parameters of the procedure. To calibrate the parameters, three test problems of about the same size as problem *p*13 were used. The parameters in column "Best" were then used for all experiments reported in this section.

The first set of tests illustrates the performance of the tabu heuristic in comparison to the branch-and-bound procedure of CPLEX. The results are presented in Table 4. The tabu search algorithm was stopped after one hour of CPU time. CPLEX was stopped after 5 hours CPU time, or when 300,000 nodes had been

Table 3 Tested and Best Parameter Settings

Parameter	Tested Values	Best
Length of tabu list for opened kernel arcs	5, 10, 15	10
Length of tabu list for closed kernel arcs	5, 10, 20	10
Number of evaluated neighbours in $N_1(\bar{y})$	25, 50, 100, all	50
Number of evaluated neighbours in $N_2(\bar{y})$	25, 50, 100, 200	50
Move to a neighbour y' of type $N_2(\bar{y})$ if $z(y') \geq z(\bar{y})$	yes, no	no
Number of iterations i between artificial arc cost updates	1, 2, 4, 8	4
Growth factor for artificial arc costs	1.1, 1.2, 1.3	1.2
Decrease factor for artificial arc costs	0.9, 0.8, 0.7	0.7

Table 4 Performance Comparisons

Problem	Tabu Solution	CPLEX MIP UBD	LP-Relaxation
*p*1	44214	*44214	41813
*p*2	23205	*23205	19449
*p*3	19208	19208	17004
*p*4	25365	25365	19004
*p*5	31501	32457	24442
*p*6	37834	39341	30387
*p*7	47624	47624	41718
*p*8	65345	65345	61270
*p*9	136809	137474	126321
*p*10	204764	210825	175934
*p*11	155361	164149	130514
*p*12	683778	696249	620556
*p*13	305933	—	276454
*p*14	188292	—	160430
*p*15	369850	—	330940
*p*16	357313	—	279848
*p*17	414617	—	328063
*p*18	338618	—	251611
*p*19	157099	—	—
*p*20	195699	—	—
*p*21	149490	—	—

examined, or when CPLEX ran out of memory. For the smaller problems (*p*1 through *p*12), CPLEX found integer valued solutions, but could prove optimality only for problems *p*1 and *p*2 (indicated with a $*$ in Table 4).

The solutions found by CPLEX were never better than the solutions found by the tabu heuristic. CPLEX solutions as good as those of tabu search were obtained only for the problems *p*1 through *p*4, *p*7 and *p*8, i.e., for the problems with the fewest binary variables.

For problems with more than 710 binary variables, the tabu heuristic was always better than CPLEX. For problems *p*13 through *p*18, CPLEX found no integer valued solutions in five hours of computing time. For problem *p*19 through *p*21, CPLEX ran out of memory (this happened before the LP-relaxation was solved).

In all test problems, the gap between the best found integer valued solution and the LP-relaxation is large. For the smaller problems (*p*1 through *p*9), the LP-relaxation is solved quickly by CPLEX. However, there is no straightforward way of achieving a good integer valued solution from an LP-relaxation solution because it is difficult to round the fractional variables so that both the node constraints and the capacity constraints remain satisfied.

The tabu search heuristic is much faster than CPLEX, which can be verified by comparing the time it takes for the tabu search to find a solution comparable to the best solution from CPLEX. For the problems *p*4, *p*7, and *p*8, for which the tabu heuristic and CPLEX found equally good solutions, the solution time for CPLEX was 7, 24, and 16 times longer, respectively. For most of the larger problems, the best solution that the tabu search heuristic could find was achieved after almost one hour, that is, just before the calculations were stopped, which indicates that the algorithm was still progressing.

Our conclusion is that the tabu search algorithm can be used to solve the proposed optimization model for problem instances of realistic size. The branch-and-bound algorithm in CPLEX can prove optimality only for very small problems, while for larger problems, the best solution found by CPLEX is inferior to solutions found by the tabu search. In addition, tabu search is much faster.

In an actual real company, trains and cars are handled as whole entities. Consequently, in principle, all variables of the model should be integer valued, since the solution is interpreted as an empty flow pattern. However, in order to simplify the solution procedure, we relax the integrality constraints on flow variables, while explicitly considering the integrality constraints on the binary variables associated to the design arcs (y variables).

In general, however, for the test problems we considered, the tabu search algorithm yielded integer valued solutions for all variables, in most steps of the search. To analyze more deeply the impact of this relaxation on the type of solutions obtained, we studied the integrality of one problem instance of each combination of number of hubs and terminals. On average, 88% of the visited solutions were integer valued and, in particular, 93% of the solutions that were recorded as "currently best solution" were integer valued. Of the noninteger valued solutions, 80% correspond to infeasible solutions (i.e., there is flow on some artificial arc). When solutions are feasible but noninteger valued, only 3% of the variables currently used (flows on open kernel arcs) are not integer valued.

Integer-valued solutions are thus obtained in most cases. Therefore, when the model is to be used to study principles of distribution performance, as in this paper, or for strategic or tactical planning, the actual procedure can be used directly. On the other hand, if the model is used for operational planning, the integrality constraints on the flow variables should not be relaxed and the applied solution method should take the integrality of the flow variables into account. In order to guarantee that the tabu search procedure results in an integer valued solution, the method can be modified to only store solutions with all integer valued variables as the "currently best solution." Since most visited solutions are integer valued anyway, this modification should not impact the performance of the method.

The second test illustrates the impact on the empty freight car distribution of the introduction of the fixed

Table 5 Impact of Car Cluster Transportation Costs

Cost per Car Cluster	Number of Car Clusters	Increase in Transportation Cost
0	309	0
50	130	774 (0.3%)
100	123	1254 (0.4%)
200	112	3459 (1.2%)
400	101	5874 (2.1%)
800	98	8694 (3.1%)

costs for transporting car clusters. Table 5 shows how the number of car clusters in the best found solutions depends on the size of the car cluster cost. (The optimal transportation cost when no costs on car clusters are considered is 280232.) The results correspond to the mean values of three test problems after one hour of calculation time, during which about 250 iterations of the tabu heuristic were performed. All three test problems have about the same size as problem *p*13 in Table 2.

The objective of introducing a cost on car clusters in the optimization model is to reduce the number of car clusters that are transported in the railway system. The results in Table 5 indicate that introducing car cluster costs into the formulation has a significant impact on solutions. As the cost for each car cluster is increased, the number of car clusters transported in the empty freight car distribution is reduced. On the other hand, the transportation cost increases. Table 5 shows the number of car clusters used in the best found solution, as well as the increase of the transportation cost compared to a solution in which costs on car clusters are not included. The results show that if the cost for sending a car cluster is increased, the transportation pattern is organized so that fewer car clusters are used while the transportation cost increases slightly. Preprocessing the problems before the tabu search algorithm was applied reduced the number of car clusters used in the solutions from 309 to 165, while the transportation cost was not increased at all.

We also conducted experiments with nonzero salvage values. No major differences in results and conclusions were found. The performance of the tabu heuristic is not influenced by the salvage values. The number of transported car clusters and the transportation cost are slightly increased as a consequence of the preferences about the transportation pattern that are stated by the salvage values, but the general conclusions stated above are still valid. The methodology we propose may thus be used in a rolling horizon planning environment.

6. Conclusion

In this paper we have formulated a model for the operational empty freight car distribution that includes explicit consideration of economies of scale. In order to capture economies of scale, we include a cost for transporting a car cluster between an origin and a destination in the railway system. The optimization model is based on a time-expanded network where the cost of a car cluster is represented as a fixed cost for assigning flow to a path. The optimization model is thus a special case of a multicommodity capacitated network design model, with the unusual feature that the capacity restrictions limit the total flow on several arcs. A tabu search metaheuristic is used to solve the model. The heuristic successfully and efficiently finds solutions to the model that are of a much better quality than those an MIP solver could find.

The solutions obtained from the model display the economies-of-scale effect, in the sense that the car movements in the solution are coordinated to reduce the number of car clusters that are transported in the railway network. For instance, in our tests, the number of car clusters transported to fulfill the supply and demand in the empty freight car distribution could be reduced by 67%, while the transportation cost increased by only 2%.

A key characteristic of the model is the definition of kernel paths and the introduction of fixed costs on such paths. This modeling approach can certainly have applications in other problem areas as well. The important aspect the model captures is the economies of scale associated with transporting more than one unit from an origin to a destination, while capacity restrictions limit the amount that can be transported. The capacity restrictions are not applicable directly to the origin-destination transports, but to sublinks that together form the transports.

A number of research issues offer interesting avenues to continue this project. The model of empty freight car distribution can be further enhanced by, for example, considering coordination between different car clusters. The mathematical formulation could be strengthened through reformulation or the addition of valid inequalities to yield, hopefully, stronger bounds. Even though the tabu heuristic used in this paper is rather simple, we achieved very promising results. A more refined tabu search strategy should permit us to further improve these results.

Acknowledgments
This work was financed by Swedish Railways (SJ/Green Cargo). Also, partial funding for this project was provided by the Natural Sciences and Engineering Council of Canada, and by the Fonds F.C.A.R. of the Province of Québec.

References

Balakrishnan, A., T. L. Magnanti, P. Mirchandani. 1997. Network design. M. Dell'Amico, F. Maffioli, S. Martello, eds. *Annotated Bibliographies in Combinatorial Optimization*. John Wiley & Sons, New York, 311–334.

Cordeau, J.-F., P. Toth, D. Vigo. 1998. A survey of optimization models for train routing and scheduling. *Transportation Sci.* **32**(4) 380–404.

Crainic, T. G. 2000. Network design in freight transportation. *Eur. J. Oper. Res.* **122**(2) 272–288.

Crainic, T. G., M. Gendreau, K. Holmberg, M. Joborn, J. T. Lundgren. 1997. A tabu heuristic for the network design problem with fixed costs on paths. Working Paper LiTH-MAT-WP-1997-01, Department of Mathematics, Linköping University, Linköping, Sweden.

Dejax, P. J., T. G. Crainic. 1987. A review of empty flows and fleet management models in freight transportation. *Transportation Sci.* **21**(4) 227–247.

Gendreau, M., A. Hertz, G. Laporte. 1994. A tabu search heuristic for the vehicle routing problem. *Management Sci.* **40** 1276–1290.

Glover, F., M. Laguna. 1997. *Tabu Search*. Kluwer Academic Publishers, Norwell, MA.

Haghani, A. E. 1989. Formulation and solution of combined train routing and makeup, and empty car distribution model. *Transportation Res.* **23B**(6) 433–452.

Holmberg, K., M. Joborn, J. T. Lundgren. 1998. Improved empty freight car distribution. *Transportation Sci.* **32**(2) 163–173.

Hughes, R., W. B. Powell. 1988. Mitigating end effects in the dynamic vehicle allocation model. *Management Sci.* **34**(7) 859–879.

ILOG. 1999. *ILOG CPLEX 6.5*. ILOG, Mountain View, CA.

Joborn, M. 2001. Optimization of empty freight car distribution in scheduled railways. Ph.D. thesis, Department of Mathematics, University of Linköping, Linköping, Sweden.

Magnanti, T. L., R. T. Wong. 1986. Network design and transportation planning: Models and algorithms. *Transportation Sci.* **18**(1) 1–55.

Minoux, M. 1986. Network synthesis and optimum network design problems: Models, solution methods and applications. *Networks* **19** 313–360.

Powell, W. B., T. A. Carvalho. 1998a. Dynamic control of logistics queueing networks for large-scale fleet management. *Transportation Sci.* **32**(2) 90–109.

Powell, W. B., T. A. Carvalho. 1998b. Real-time optimization of containers and flatcars for intermodal operations. *Transportation Sci.* **32**(2) 110–126.

Powell, W. B., T. A. Carvalho, G. A. Godfrey, H. P. Simaõ. 1995. Dynamic fleet management as a logistics queueing network. *Ann. Oper. Res.* **61** 165–188.

Turnquist, M. A. 1994. Economies of scale and network optimization for empty car distribution. Technical report, Cornell University, Ithaca, NY.