CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

A Brief Overview of Intermodal Transportation

Tolga Bektas
Teodor Gabriel Crainic

January 2007

CIRRELT-2007-03

Bureaux de Montréal:
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone: 514 343-7575
Télécopie: 514 343-7121

Bureaux de Québec:
Université Laval
Pavillon Palasis-Prince, local 2642
Québec (Québec)
Canada G1K 7P4
Téléphone: 418 656-2073
Télécopie: 418 656-2624

www.cirrelt.ca

Université Laval · UQÀM · HEC Montréal · École Polytechnique Montréal · Université de Montréal

A Brief Overview of Intermodal Transportation

Tolga Bektas[1], Teodor Gabriel Crainic[1,*]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7 and Chaire de recherche industrielle du CRSNG en management logistique, ESG, Université du Québec à Montréal, C.P. 8888, succursale Centre-ville, Montréal, Canada H3C 3P8

Abstract. This paper focuses on Intermodal Freight Transportation broadly defined as a chain made up of several transportation modes that are more or less coordinated and interact in intermodal terminals to ensure door-to-door service. The goal of the chapter is to present intermodal transportation from both the supplier and the carrier perspectives, and identify important issues and challenges in designing, planning, and operating intermodal transportation networks, focusing on modeling and the contributions of operations research to the field.

Keywords. Intermodal transportation, freight transportation, operations research

Acknowledgements. Partial funding for this project has been provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) through its Discovery Grants and Chairs and Faculty Support programs.

---

* Corresponding author: Teodor-Gabriel.Crainic@cirrelt.ca

Dépôt légal – Bibliothèque nationale du Québec,
Bibliothèque nationale du Canada, 2007

I. Introduction

In today's world, intermodal transportation forms the backbone of world trade. Contrary to conventional transportation systems in which different modes of transportation operate in an independent manner, intermodal transportation aims at integrating various modes and services of transportation to improve the efficiency of the whole distribution process. Parallel to the growth in the amount of transported freight and the changing requirements of integrated value (supply) chains, intermodal transportation exhibits significant growth. According to the U.S. Department of Transportation (2006), the value of the multimodal shipments, including parcel, postal service, courier, truck-and-rail, truck-and-water, and rail-and-water increased from about $662 billion to about $1.1 trillion in a period of nine years (1993 to 2003).

Major players in intermodal transportation networks are *shipper*s, who generate the demand for transportation, *carrier*s, who supply the transportation services for moving the demand, and the intermodal network itself composed of multimodal services and terminals. The interactions of these players and their individual behavior, expectations, and often conflicting requirements determine the performance of intermodal transportation systems. The goal of this chapter is therefore to be informative on intermodal transportation, from both the supplier and the carrier perspective, identify important issues and challenges in designing and operating intermodal transportation networks, and point out major operations research contributions to the field. A more in-depth discussion of these topics may be found in, for example, Crainic and Kim (2007), Macharis and Bontekoning (2004), and Sussman (2000).

The chapter is structured as follows. Section II presents the basics on intermodal transportation, with an emphasis on its foremost components: containers, carriers, and shippers. We then discuss, in Sections III and IV, respectively, the major issues and challenges of intermodalism from the shippers and carriers perspective. Section V provides a brief description of intermodal terminals and the operations performed therein. Section VI is dedicated to the case of rail intermodal transportation, as an illustration of the main discussion.

II. Intermodal Transportation

Many transportation systems are *multimodal*, that is, the infrastructure supports various transportation modes, such as truck, rail, air, and ocean/river navigation, carriers operating and offering transportation services on these modes. Then, broadly defined, *intermodal* *transportation* refers to the transportation of people or freight from their origin to their destination by a sequence of at least two transportation modes. Transfers from one mode to the other are performed at *intermodal terminals*, which may be a sea port or an in-land terminal, e.g., rail yards, river ports, airports, etc. Although both people and freight can be transported using an intermodal chain, in this chapter, we concentrate on the latter.

The fundamental idea of intermodal transportation is to consolidate loads for efficient long-haul transportation (e.g., by rail or large ocean vessels), while taking advantage of the efficiency of local pick-up and delivery operations by truck. This explains the importance of container-based transportation. Freight intermodal transportation is indeed often equated to moving containers over long distances through multimodal chains. Intermodal transportation is not restricted, however, to containers and intercontinental exchanges. For instance, the transportation of express and regular mail is intermodal, involving air and land long-haul transportation by rail or truck, as well as local pick up and delivery operations by truck (Crainic and Kim 2007). In this paper, we focus on container-based transportation.

An intermodal transportation chain is illustrated in Figure 1. In this example, loaded containers leave the shipper's facilities by truck to a rail yard, where they are consolidated into a train and sent to another rail yard. Trucks are again used to transport the containers from this rail yard to the sea container terminal. This last operation may not be necessary if the sea container terminal has an interface to the rail network, in which case freight is transferred directly from one mode to the other. Containers are then transported to a port on another continent by ocean shipping, from where they leave by either trucking or rail (or both) to their destinations.

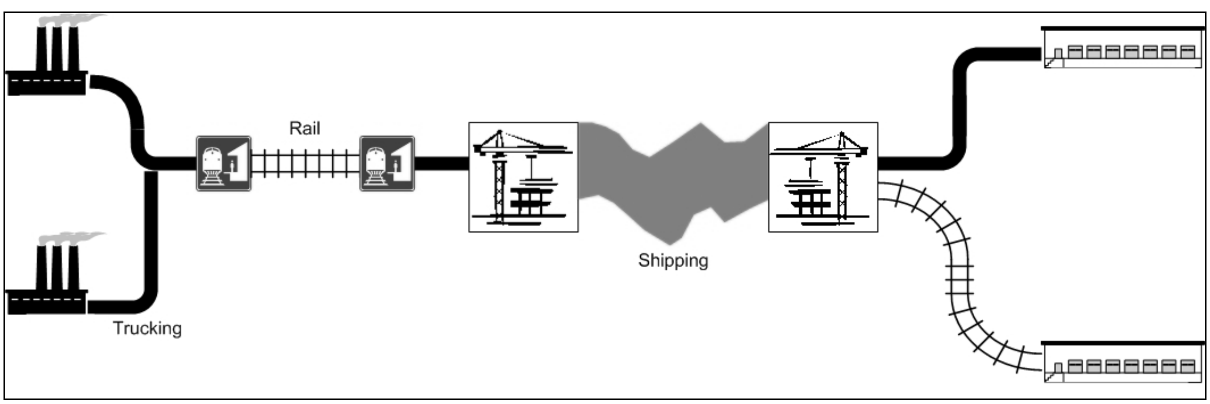


Figure 1. An intermodal transportation network

II.1 Containers

A *container*, as defined by the European Conference of Ministers of Transport (2001), is a "generic term for a box to carry freight, strong enough for repeated use, usually stackable and fitted with devices for transfer between modes". The fact that the standards on container dimensions were established very early also explains its popularity. A standard container is the 20-foot box, which is 20 feet long, 8'6" feet high and 8 feet wide. This is refereed to as a *Twenty-foot Equivalent Unit* (*TEU*). However, the widely used container size is the 40-foot box (a number of longer boxes are sometimes used for internal transport in North America). Containers are either made of steel or aluminum, the former being used for maritime transport and the latter for domestic transport.

Intermodal transportation relies heavily on containerization due to its numerous advantages. First, containerization offers safety by significantly reducing loss and damage, since the contents of a container cannot easily be modified unless except at origin or destination. It is worth mentioning in this respect that the safety level of container transportation is currently being significantly increased by electronic sealing and monitoring to address preoccupations with terrorist treats, illegal immigration, and smuggling. Second, due to its standard structure, transfer operations at terminals are fast and performed with a minimal amount of effort. This results in reduced cargo handling, and thus a speed-up of operations not only at the terminals, but through the whole transport chain. Third, containers are flexible enough to enable the transport of products of various types and dimensions. Fourth, containerization enables a better management of the transported goods. Due to these reasons, the use of containers significantly decreases transport costs.

Containerization has had a noteworthy impact on both land transportation and the way terminals are structured. An example for the former can be seen in rail transportation, where special services have been established by North-American railways, enabling container transportation by long, double-stack trains. As for the latter, ports and container terminals have either been built or undergone major revisions to accommodate continuously larger container ships and efficiently perform the loading, unloading, and transfer operations. Container terminal equipment and operating procedures are continuously enhanced to improve productivity and compete, in terms of cost and time, with the other ports in attracting ocean shipping lines.

II.2 Carriers

In an intermodal chain, carriers may either provide a *customized* service, where the vehicle (or convoy) is dedicated exclusively to a particular customer, or operate on the basis of *consolidation*, where each vehicle moves freight for different customers with possibly different origins and destinations.

Full-load trucking is a classic example of customized transportation. Upon the call of a customer, the truck is assigned to the task by the dispatcher. The truck then travels to the customer location, is loaded, and then moves to the destination, where it is unloaded. Following this, the driver is either assigned a new task by the dispatcher, kept waiting until a new demand appears in the near future, or repositioned to a location where a load exists or is expected to be available about the arrival time. The advantages of full-load trucking come from its flexibility in adapting to a highly dynamic environment and uncertain future demands, offering reliability in service and low tariffs compared to other modes of transportation. The fill efficiency of full-load trucking is achieved through the implementation of *resource management and allocation strategies* that seek to make the best use of the available resources, while maximizing the volume of demand satisfied and the associated profits (Powell and Topaloglu, 2005, Powell, Bouzaïene-Ayari, and Simaõ, 2007). Customized services are also offered, for example, by charted sea or river vessels and planes.

In many cases, however, trade-offs between volume and frequency of shipping, along with the cost of transportation, render customized services impractical. Consolidation, in such situations, turn out to be an attractive alternative. Freight consolidation transportation is performed by Less-Than-Truckload (LTL) motor carriers, railways, ocean shipping lines, regular and express postal services, etc. A consolidation transportation system is structured as a *hub-and-spoke* network, where shipments for a number of origin-destination points may be transferred via intermediate consolidation facilities, or *hub*s, such as airports, seaport container terminals, rail yards, truck break-bulk terminals, and intermodal platforms. An example of such a network with three hubs and seven regional terminals (origin and destination points for demand) is illustrated in Figure 2. In hub-and-spoke networks, low-volume demands are first moved from their origins to a hub where the traffic is sorted (classified) and grouped (consolidated). The aggregated traffic is then moved in between hubs by high frequency and high capacity services. Loads are then transferred to their destination points from the hubs by lower frequency services often utilizing smaller vehicles. When the level of demand is sufficiently high, direct services may be run between a hub and a regional terminal. Although a hub-and-spoke network structure results in a more efficient utilization of resources and lower costs for shippers, it also incurs a higher amount of delays and a lower reliability due to longer routes and the additional operations performed at terminals. The planning methodologies evoked at Section 4 aim to address these issues.

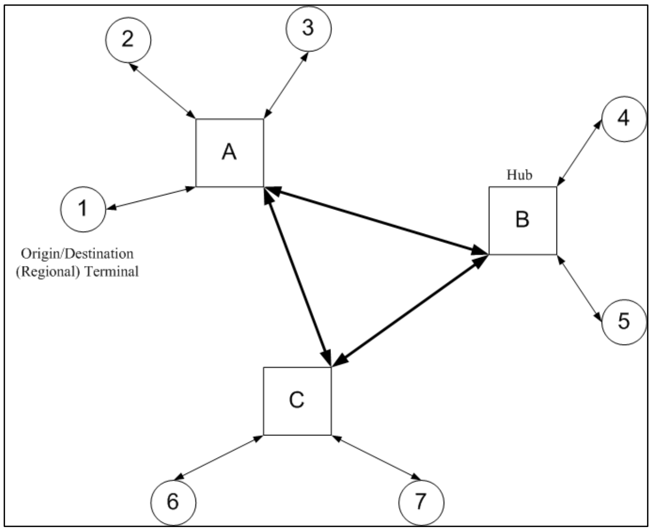


Figure 2. A *hub-and-spoke* network

Land consolidation transportation services are offered by *LTL motor carriers* and *railways*. The flexibility, high frequency, and low cost of trucking transportation resulted in a high market share of freight transportation being captured by this mode. This situation, which may be observed world-wide, resulted in very large truck flows and road congestion, and contributes significantly to the high level of pollutant emissions attributed to the transportation sector. The trend is slowly changing, however. On the one hand, recent policy measures, particularly in the European community, target the mode change from road-based to intermodal. On the other hand, the continuous and significant increase in container-based international traffic, which generates large flows of containers that need to be moved over long distances, favors rail (and, in a smaller measure, river) based transportation.

Railways have risen to the challenge by proposing new types of services and enhanced performances. Thus, North-American railways have created intermodal subdivisions which operate so-called “land-bridges” providing efficient container transportation by long, double-stack trains between the East and the West coasts and between these ports and the industrial core of the continent (so-called "mini" land-bridges). New container and trailer-dedicated shuttle-train networks are being created within the European Community. These initiatives have succeeded. Thus, intermodal transportation is the fastest growing part of railroad traffic. Thus, for example, in a period of only 2 years, from 2003 to 2005, the amount of rail intermodal traffic in the U.S. grew from 7.33 million to 8.71 million containers (Association of American Railroads, 2006). Note that the 8.71 million containers transported by rail in U.S. in 2005 represent about 75% of the total intermodal units moved in that year, the remaining 25% being trailers. In Canada, there is a similar trend in the number of intermodal carloads, which grew by 336 000 or 77.5% during the past decade. This number represents 26.3% of the industry's overall growth in originated carloads during this period (The Railway Association of Canada, 2006). Section 6 discusses rail intermodal transportation in some more depth.

Maritime and *air-based* modes are used for the intercontinental legs of intermodal transportation. Heavily container-based, the former provide the backbone of non-bulk international trade. Efficiency reasons result in continuously larger container vessels being commissioned for inter-continental movements. The operation of such ships, which cannot pass through the Panama Canal, should not stop too frequently, and cannot even berth in most ports, has had a number of important consequences. In particular, maritime and land transportation routes have been modified through, for example, the creation of the North-American land-bridges and the introduction of a "new" link into the intermodal chain: super-ships stop at a small number of major seaports and containers are transferred to smaller ships for distribution to various smaller ports. For maritime transportation, variations in travel times are larger, and travel and loading/unloading times are longer, compared to that of most land-based trips. As for the air mode, although being increasingly used for intermodal transport, its relatively higher costs still make it interesting mostly for high-value or urgent deliveries.

All consolidation-based transportation modes involved in intermodal transportation must provide efficient, reliable, and cost-effective services. Comparatively, railways face the biggest challenge, however, in competing with trucking to provide shippers the level and quality of service they require for their land long-haul transportation needs. Sections IV and VI discuss these issues.

II. 3 Shippers

Shippers generate the demand for transportation. Defining its logistics strategy represents a complex decision process, and the choice of the transport mode is only a part of this whole strategy. This process is generally assumed to have a three-level decision structure, composed of long, medium and short term decision (Bolis and Maggi, 2003). In the long run, shippers define their logistics strategies in terms of their customer network and production. Medium-term plans include decisions as to inventories at production, warehousing, and distribution facilities, frequency and amount of shipping, flexibility of service, etc. Finally, the shipper decides, at the short-term level, the attributes of the services required for its shipments, such as maximum rates, transport time, reliability and safety.

When such decisions are made, shippers consider the availability and the characteristics of the services offered on the market by carriers and intermediaries such as freight brokers and third-party logistics providers. These decisions are based on several factors detailed in the following section. It’s worth mentioning, however, that in many cases shipper decisions have a greater importance for the outcome of the service rather than the way it is delivered.

III. Shipper Perspectives on Intermodal Transportation

Although intermodal networks are formed as combinations of individual transportation modes and transfer facilities, a shipper perceives it as a single integrated service. Shippers therefore expect intermodal services to behave similarly to unimodal services, especially in terms of speed, reliability, and availability.

The shipper's decision to use a particular transportation mode is generally based on several criteria. A number of studies mostly based on surveys and data analyses have been conducted to identify the specific service characteristics often deemed important in the shipper decision process. McGinnis (1990) identified six factors that affect the shipper’s decision in choosing a specific transportation mode, namely 1) freight rates (including cost and charges), 2) reliability (delivery time), 3) transit times (time-in-transit, speed, delivery time), 4) over, short and damaged shipments (loss, damage, claims processing, and tracing), 5) shipper market considerations (customer service, user satisfaction, market competitiveness, market influences), and 6) carrier considerations (availability, capability, reputation, special equipment). Amongst these, shippers were observed to place more emphasis on overall service than cost, although freight rates still had a significant importance. Five years later, in an update to McGinnis' study, Murphy and Hall (1995) observed that reliability, as opposed to cost or any other factor, had more influence on the U.S. shipper decisions.

As important as the shipper's *choice* of transportation modes may be, shipper *perception* of modes and services is believed to have a higher impact on the overall decision process. Quoting from Evers *et al.* (1996), “shippers decide on the mode of transportation and specific carrier only after they have formed perceptions of the alternative services. They compare their perceptions, and possibly other information, with criteria they have developed. Then, they use a decision-making process to choose the transportation method that best meets their established criteria”. Hence, shipper perception is a central input component to the decision making process in mode selection.

In a study conducted to identify the determinants of shipper perceptions of modes, Evers *et al*. (1996) proposed six factors: 1) timeliness, 2) availability, 3) firm contract, 4) suitability, 5) restitution, and 6) cost, the first two being observed to have a greater effect than the others. Thus, the more a carrier focuses on improving shipper’s perceptions of these six factors, the more likely it is for it to be used. Extending their analysis to an individual carrier level, Evers and Johnson (2000) found that the future collaboration of a shipper with an intermodal service is affected by the shipper’s satisfaction with, and ability to replace, the carrier. They identified that determinants of shipper’s perceptions at the carrier level were communication, quality of customer service, consistent delivery, transit times, and competitive rates, with the first two factors being the most important drivers of the overall perception. Although these results do confirm what is expected of a carrier, it is interesting to see that the perception of modes may be different from that of carriers. More specifically, the two chief factors influencing the overall perception for modes, namely *timeliness* and *availability*, are replaced by *communication* and *customer service* for the individual carriers.

Studies performed in the early 1990’s revealed that shippers have varying perceptions of alternative transportation modes. In general, shippers were found to perceive truck transportation best, followed by intermodal (rail-truck) transportation, and rail transportation. Prior research indicates that perceptions of railroad service have improved since deregulation in 1980, both in terms of rates and service. More recent research shows that *price*, *time*, and *reliability* are important factors in the decision process, but, *frequency* and *flexibility* also emerge as significant decision factors when firms operate in a JIT context. Empirical evidence exists, on the other hand, as to the fact that, in Europe, rail has no acceptance problems but that of service quality (Bolis and Maggi, 2003).

Shipper perceptions, influenced by past experiences, expectations, common knowledge, carrier advertising, modal image, and misinformation (Evers *et al*., 1996), may not always reflect the real situation. After all, a shipper's perception of a mode or an individual carrier is a tentative view. This, however, may be altered through marketing efforts by the carrier. Carriers thus should strive to ensure that the shipper's perception is in line with reality as much as possible as such efforts can be beneficial, in terms of increased usage of their service. Failure to put forth such efforts may easily place the carrier at a disadvantage with respect to its competitors.

IV. Carrier Perspective on Intermodal Transportation

Carriers face a number of issues and challenges in providing an efficient and cost effective service to the customer, which may be examined according to classical categorization of planning decisions, namely *strategic* (long term), *tactical* (medium term), and *operational* (short term) level of planning and management of operations. We briefly address these issues in this section by focusing on consolidation-based carrier cases. This choice is motivated by the complexity of the planning issues in this context and on the fact that intermodal transportation forms a consolidation-based system. For a more in-depth treatment, one may consult the reviews of Christiansen *et al.* (2005, 2007), Cordeau *et al.* (1998), and Crainic (2003), Crainic and Kim (2007), Crainic and Laporte (1997).

IV.1 System Design

At the strategic level, the carrier is concerned with the design of the physical infrastructure network involving decisions as to the number and location of terminals (e.g., consolidation terminals, rail yards, intermodal platforms), the type and quantity of equipment (e.g., cranes) that will be installed at each facility, the type of lines or capacity to add or abandon, the customer zones to serve directly, etc. The term *system design* encompasses issues pertinent to strategic level decisions. It is often the case that such decisions are made by evaluating alternatives using network models for tactical or operational planning of transportation activities. When specific models are developed for strategic planning issues, these usually take the form of static and deterministic location formulations addressing issues related to the location of consolidation or hub terminals and the routing of demand from its origin to its destination terminals.

To illustrate the methodological approaches proposed to address strategic planning issues for consolidation-based carriers, we examine the issue of determining the sub structure of such a system. This problem consists in determining the locations of the hubs on a given network, the assignment of local terminals to the hubs that are established, and the routing of loads of each demand through the resulting network. In its most basic form, the problem is addressed by a multi-commodity hub-location formulation, assuming that all traffic passes through two hubs on its route from its origin to its destination. In this simple formulation, it is assumed that there can be no direct transport between non-hub terminals, based on the hypothesis that inter-hub transportation is more efficient due to consolidation. Furthermore one also assumes that there are neither capacity restrictions on hubs, nor fixed costs associated with establishing a link between a regional and a consolidation terminal.

The problem is modeled on a directed network $G = (N, A)$, with N as the set of nodes (or vertices) and A is the set of arcs (or links). Nodes are identified as origins (set O), destination (set D), and hubs or consolidation nodes *(set H*; sets O, N, and H are not necessarily disjoint). The set of commodities (types of containers) that move through the network are represented by the set P. The amount of commodity $p \in P$ to be transported from origin terminal $i \in O$ to destination terminal $k \in D$ is denoted by d_{ik}^{p}. There are three decision variables. y_j is a binary variable equal to 1 if a consolidation terminal is located at site j and 0, otherwise. y_{ij} is a binary variable equal to 1 if terminal i is linked to hub j and 0, otherwise. Finally, variable x_{ijlk}^{p} denotes the amount of flow for commodity p with origin i, destination k, passing through terminals j and l in the given order. The following formulation then solves the aforementioned network design problem with exactly M hubs are installed in the network, where c_{ij}^{p}, c_{lk}^{p}, and c_{jl}^{p} stand for the unit transportation costs between origin terminals and hubs, hubs and destination terminals, and inter-hubs, respectively.

$$\text{Minimize} \quad \sum_{p \in P} \left\{ \sum_{i \in O} \sum_{j \in H} c_{ij}^{p} y_{ij} \left(\sum_{l \in H} \sum_{k \in D} x_{ijlk}^{p} \right) + \sum_{l \in H} \sum_{k \in D} c_{lk}^{p} y_{lk} \left(\sum_{i \in O} \sum_{j \in H} x_{ijlk}^{p} \right) + \sum_{j \in H} \sum_{l \in H} c_{jl}^{p} y_{ij} y_{kl} \left(\sum_{i \in O} \sum_{k \in D} x_{ijlk}^{p} \right) \right\}$$

$$\text{subject to} \quad \sum_{j \in H} y_j = M, \tag{4.1}$$

$$\sum_{j \in H} \sum_{l \in H} x_{ijlk}^{p} = d_{ik}^{p} \qquad i \in O, k \in D, p \in P, \tag{4.2}$$

$$y_{ij} \leq y_j \qquad i \in O, j \in H, \qquad (4.3)$$

$$y_{kl} \leq y_l \qquad k \in D, l \in H, \qquad (4.4)$$

$$x_{ijlk}^p \leq d_{ik}^p y_j \qquad i \in O, k \in D, j, l \in H, p \in P, \qquad (4.5)$$

$$x_{ijlk}^p \leq d_{ik}^p y_l \qquad i \in O, k \in D, j, l \in H, p \in P, \qquad (4.6)$$

$$y_j \in \{0,1\} \qquad j \in H, \qquad (4.7)$$

$$y_{ij} \in \{0,1\} \qquad i \in O, j \in H, \qquad (4.8)$$

$$x_{ijlk}^p \geq 0 \qquad i \in O, k \in D, j, l \in H, p \in P. \qquad (4.9)$$

The objective function of the formulation minimizes the total transportation cost of the system. Constraints (4.1) state that exactly *M* hubs should be located in the network. To ensure that the demands are satisfied, constraints (4.2) are used. Constraints (4.3) and (4.4) guarantee that a terminal is assigned to a hub only if the hub is established. Constraints (4.6) and (4.7) serve a similar purpose in terms of routing the flows using only selected hubs.

The previous quadratic formulation was first introduced by O'Kelly (1987), while, Campbell (1994) proposed its first linearization. Both formulations are difficult, as are more complex models that include capacities, fixed costs to open facilities, link terminals to hubs, or establish transportation connections, more complex load routing patterns, etc. Consequently, although several contributions have been made relative to the analysis of hub location problems and the development of solution procedures, this is still an active and interesting field of research (see, e.g., the surveys of Campbell *et al.,* 2000 and Ebery *et al*., 2000).

IV.2 Service Network Design

Designing the service network of a consolidation-based carrier refers to constructing the transportation (or load) plan to serve the demand, while at the same time operating the system in an efficient and profitable manner. These plans are built given an existing physical infrastructure and a fixed amount of resources, as determined during the system design phase.

Service network design is concerned with the planning of operations related to the selection, routing, and scheduling of services, the consolidation of activities at terminals, and the routing of freight of each particular demand through the physical and service network of the company. These activities are a part of the tactical planning at a system-wide level. The two main types of decisions that are considered in service network design are to determine the service network and the routing of demand. The former refers to selecting the routes, characterized by origin-destination nodes, intermediate stops and the physical route, and attributes, such as the *frequency* or the *schedule,* of each service. The latter is concerned with the *itineraries* that specify how to move the flow of each demand, including the services and terminals used, the operations performed in these terminals, etc. Although minimization of total operating cost is the main criterion of the service network design objective, improving the quality of service measured by its speed, flexibility, and reliability is increasingly being considered as an additional component of this goal. Service performance measures modeled, in most cases, by delays incurred by freight and vehicles or by the respect of predefined performance targets, are then added to the objective function of the network optimization formulation. The resulting generalized-cost function thus captures the tradeoffs between operating costs and service quality. For further details, the reader is referred to Crainic (2000) and Crainic and Kim (2007).

Formulations for service network design either assume that the demand does not vary during the planning period (static formulations) or explicitly consider the distribution of demand as well as the service departures and the movements of services and loads in time (time-dependent formulations). In both cases, however, modeling efforts take the form of *deterministic*, *fixed cost*, *capacitated*, *multicommodity network design* formulations. To illustrate such approaches, we provide here the multimodal multicommodity path-flow service network design modeling framework proposed by Crainic and Rousseau (1986; see Powell and Sheffi, 1989 for a complementary formulation) for the static case.

In this formulation, the service network, defined on a graph $G = (N, A)$ representing the physical infrastructure of the system, specifies the transportation services that could be offered. Each service $s \in S$ is characterized by its 1) mode, which may represent either a specific transportation mode (e.g., rail and truck services may belong to the same service network), or a particular combination of traction and service type; 2) route, defined as a path in A, from its origin terminal to its destination terminal, with intermediary terminals where the service stops and work may be performed; 3) capacity, which may be measured in load weight or volume, number of containers, number of vehicles (when convoys are used to move several vehicles simultaneously), or a combination thereof; 4) service class that indicates characteristics such as preferred traffic or restrictions, speed and priority, etc. To design the service network thus means to decide what service to include in the transportation plan such that the demands and the objectives of the carrier are satisfied. When a service is operated repeatedly during the planning period, the design must also determine the frequency of each service.

A commodity $p \in P$ is defined as a triplet (origin, destination, type of product or vehicle) and traffic moves according to itineraries. An itinerary $l \in L_p$ for commodity p specifies the service path used to move (part of) the corresponding demand: the origin and destination terminals, the intermediary terminals where operations (e.g., consolidation and transfer) are to be performed, and the sequence of services between each pair of consecutive terminals where work is performed. The demand for product p is denoted by d_p. Flow routing decisions are then represented by decision variables h_l^p indicating the volume of product p moved by using its itinerary $l \in L_p$. Service frequency decision variables y_s, $s \in S$, define the level of service offered, i.e., how often each service is run during the planning period. Let $F_s(y)$ denote the total cost of operating service s, and $C_l^p(y,h)$ denote the total cost of moving (part of) product p demand by using its itinerary l. Further, a penalty term $\theta(y,h)$ is included in the objective function capturing various relations and restrictions, such as the limited service or infrastructure capacity. The following formulation can then be used for the service network design problem:

$$
\begin{aligned}
\text{Minimize} \quad & \sum_{s \in S} F_s(y) + \sum_{p \in P} \sum_{l \in L_p} C_l^p(y,h) + \theta(y,h) & \\
\text{subject to} \quad & \sum_{l \in L_p} h_l^p = d_p & p \in P, \\
& (y_s, x_l^p) \in \mathrm{X} & s \in S,\ l \in L_p,\ p \in P, \\
& y_s \geq 0 \text{ and integer} & s \in S, \\
& h_l^p \geq 0 & l \in L_p,\ p \in P,
\end{aligned}
$$

where $(y_s, x_l^p) \in \mathrm{X}$ stand for the classical linking constraints (i.e., no flow may use an unselected service) as well as additional constraints reflecting particular characteristics, requirements, and policies of the particular firm (e.g., particular routing or load-to-service assignment rules). The objective function of this formulation describes a generic cost structure, flexible enough to accommodate various productivity measures related to terminal and transportation operations. As an example, one may consider service capacity restrictions as utilization targets, which may be allowed to be violated at the expense of additional penalty costs. The last component of the objective function, albeit in a nonlinear form, may be used to model such a situation.

The network design, in general, and service network design, in particular, problems are difficult and transportation applications tend to be of large dimensions with complicating additional constraints. A number of important contributions to both methodological developments and applications have been proposed, and are reviewed in the references indicated at the beginning of this section. Many challenges still exist, however, and make for a rich research and development field.

IV.3 Operational Planning

The purpose of operational level planning is to ensure that the system operates according to plan, demand is satisfied, and the resources of the carrier are efficiently used. Most methodologies aimed at carrier operational-planning issues explicitly consider the time dimension and account for the dynamics and stochasticity inherent in the system and its environment, some having to be solved in real or near-real time (e.g., dynamic resource allocation).

Main operational-level planning issues relate to *empty vehicle distribution and repositioning*, also sometimes called fleet management, *crew scheduling,* including the assignment of crews to vehicles and convoys, and *allocation of resources,* such as the dynamic allocation empty vehicles to terminals, motive power to services, crews to movements or services, loads to driver-truck combinations, *routing of vehicles* for pick up and delivery activities, and the real-time adjustment of services, routes, and plans following modifications in demands, infrastructure conditions (e.g., breakdowns, accidents or congestion), weather conditions, and so on.

We will not attempt here to review this field of research, which has been studied extensively for various modes of transport. We refer the reader to the surveys by Christiansen *et al.* (2005, 2007), Cordeau *et al*. (1998), Crainic and Laporte (1997), Crainic and Kim (2007), Powell *et al*. (2005, 2007), Powell and Topaloglu (2005), Toth and Vigo (2002) and the references therein, for details on these planning issues. It's worth mentioning, however, that few efforts were dedicated to container fleet management issues: Crainic *et al.* (1993) proposed a series of deterministic and stochastic models for the allocation and management of a heterogeneous fleet of containers where loaded movements are exogenously accepted; Cheung and Chen (1998) focused on the single-commodity container allocation problem for operators of regular ocean navigation lines and proposed a two-stage stochastic model; while Powell and Carvalho (1998; see also Powell *et al*., 2007) and Powell and Topaloglu, 2005) addressed the problem of the combined optimization of containers and flatcars for rail intermodal operations using an adaptive dynamic stochastic programming approach. Significant more research is thus required in this field.

V. Intermodal Terminals

Intermodal terminals may belong to a given carrier, rail yards, for example, or be operated independently on behalf of public or private firms (e.g., air and sea and river ports). The main role of these facilities is to provide the space and equipment to load and unload (and, eventually, store) vehicles of various modes for a seamless transfer of loads between modes. When containerized traffic is of concern, the operations performed are restricted to the handling of the containers and not the cargo they contain. Terminal operations may also include cargo and vehicle sorting and consolidation, convoy make up and break down, and vehicle transfer between services. Some terminals, sea ports and airports, in particular, also provide the first line of customs, security, and immigration control for a country. Thus, for example, North American sea port terminals are being equipped with container scanning facilities for enhanced control and security. Terminals thus form perhaps the most critical components of the entire intermodal transportation chain, as the efficiency of the latter highly depends on the speed and reliability of the operations performed in the former. Avoiding unplanned delays and the formation of load or vehicle bottlenecks is one of the major goals in operating intermodal terminals. Given the space limitations of this chapter, we only indicate the major classes of operations and issues related to intermodal terminals. For a more in-depth study of these issues, see the reviews of Crainic and Kim (2007), Günther and Kim (2005), and Steenken (2004).

The intermodal transfer of containers between truck and rail take place at *rail yards*. When containers arrive at a rail yard by truck, they are either directly transferred to a rail car or, more frequently, are stacked in a waiting area. Containers are then picked up from the waiting area and loaded unto rail cars that will be grouped into blocks and trains. When containers arrive by train to the terminal, they are transferred to trucks using the reverse operations.

Major operations in rail yards are: *classification* (sorting of rail cars), *blocking* (consolidation of rail cars into blocks), and *train make-up* (forming of blocks to trains). A significant amount of research exists on planning these operations. In most cases, there are no differences between intermodal and regular rail traffic with respect to blocking and train make up planning and operations, even when particular terminals are dedicated to handling intermodal traffic. See Bostel and Dejax (1998) for models that target planned terminals dedicated to the transfer of containers among intermodal shuttle trains.

A *container port terminal* provides transfer facilities for containers between sea vessels and land transportation modes, in particular, truck and rail. Such terminals are composed of three areas. The *sea-side* area includes the quays where ships berth and the quay-cranes that facilitate the loading and unloading of containers into and from ships. The truck and train receiving gates are located on the *land-side area*, which constitutes an interface between the land and sea transportation systems. Rail cars are loaded and unloaded in this area. Finally, there is the *yard area*, reserved mostly for stacking loaded and empty containers, and for loading and unloading the trucks.

Operations at a container port terminal can be partitioned into three classes: The first class consists of operations that deal with the berthing, loading, and unloading container ships. Following the arrival of a ship at the terminal, it is assigned a berth and a number of quay cranes. A number of planning problems arise here, such as determining the berthing time and position of a container ship at a given quay (*berth scheduling*), deciding on the vessel that each quay crane will serve and the associated service time (*quay-crane allocation*), and establishing the sequence of unloading and loading containers, as well as the precise position of each container that is to be loaded into the ship (*stowage sequencing*). The operations belonging to the second class are associated with receiving/delivery trucks and trains from/to the land-side. Containers arrive at the gate of the terminal, either by truck or train. Following inspection, the trucks are directed to the yard area where containers are unloaded and stacked. Trucks then either leave empty or pick up a new container. Empty trucks also call at the terminal to pick up containers. When containers arrive by rail, they are transferred, via a gantry crane, onto a transporter, which moves them to the designated area. The same transporter-gantry crane combination is used to load containers on departing trains. (Several variations exist according to the layout and operation mode of the terminal; the fundamental planning issues are still the same, however.) The last class of operations is concerned with container handling and storage operations in the yard. Determining the storage locations of the containers in the yard, either individually or as a group, is referred to as the *space-allocation* problem. This issue is a critical planning component as the way containers are located in the yard greatly affects the turn-around time of ships and land vehicles.

Decisions regarding allocation and dispatching of yard cranes and transporters, often performed in real or quasi-real time, complete the yard-related set of issues. The references indicated at the beginning of this section detail these issues, present the main methodologies proposed to address them, and identify interesting research perspectives and challenges.

VI. Case Study: New Rail Intermodal Services

The performance of intermodal transportation directly depends on the performance of the key individual elements of the chain, navigation companies, rail and motor carriers, ports, etc., as well as on the quality of interactions between them regarding operations, information, and decisions. The Intelligent Transportation Systems and Internet-fueled electronic business technologies provide the framework to address the latter challenges. Regarding the former, carriers and terminals, on their own or in collaboration, strive to continuously improve their performance. The rail industry is no exception. Indeed, compared to the other modes, railways face significant challenges in being a part of intermodal networks and competing intensively with trucking in offering customers timely, flexible, and long-haul transportation services.

The traditional operational policies of railroads were based on long-term contracts, providing “sure” high volumes of (very often bulk) freight to move. Cost per ton/mile (or km) was the main performance measure, with rather little attention paid to delivery times. Consequently, rail services in North America, and mostly everywhere else in the world, were organized around loose schedules, indicative cut-off times for customers, “go-when-full” operating policies, and significant marshalling (classification and consolidation of cars) activities in yards. This resulted in rather long and unreliable trip times that generated both inefficient asset utilization and loss of market share. This was not appropriate for the requirements of intermodal transportation and the North American rail industry responded through (Crainic *et al.*, 2006):

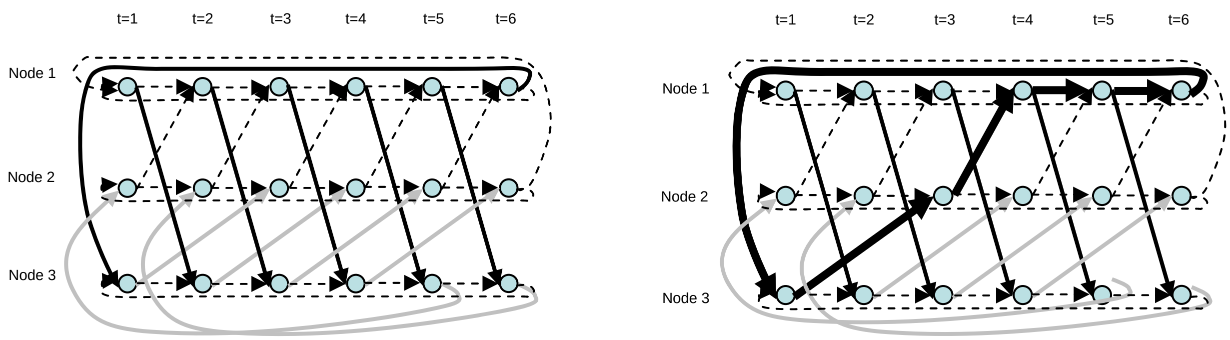
1. A significant re-structuring of the industry through a series of mergers, acquisitions, and alliances which, although far from being over, has already drastically reduced the number of companies resulting in a restricted number of major players.
2. The creation of separate divisions to address the needs of intermodal traffic, operating dedicated fleets of cars and engines, and marshalling facilities (even when located within regular yards). Double-stack convoys have created the land-bridges that ensure an efficient container movement across North America.

3. An evolution towards planned and scheduled modes of operation and the introduction of booking systems and full-asset-utilization operating policies.

Most Western Europe railways have for a long time now operated their freight trains according to strict schedules (similarly to their passenger trains). This facilitated both the interaction of passenger and freight trains and the quality of service offered to customers. Particular infrastructure (e.g., low overpasses) and territory (short inter-station distances) make for shorter trains than in North America and forbid double-stack trains. Booking systems are, however, being implemented and full-asset-utilization and revenue management operating policies are being contemplated. Moreover, shuttle-service networks are being implemented in several regions of the European Union to address the requirements of intermodal traffic (e.g., Andersen and Christiansen, 2006 and Petersen and Crainic, 2007).

Booking systems bring intermodal rail freight services closer to the usual mode of operation of passenger services by any regular mode of transportation, train, bus, or air. In this context, each class of customers or origin-destination market has a certain space allocated on the train and customers are required to call in advance and reserve the space they require. The process may be phone or Internet based but is generally automatic, even though some negotiations may occur when the train requested by the customer is no longer available. This new approach to operating intermodal rail services brings advantages for the carrier, in terms of operating costs and asset utilization, and the customers (once they get used to the new operating mode) in terms of increased reliability, regular and predictable service and, eventually, better price.

A full-asset-utilization operation policy generally corresponds to operating regular and cyclically-scheduled services with fixed composition. In other words, given a specific frequency (daily or every *x* days), each service occurrence operates a train of the same capacity (length, number of cars, tonnage) and the same number and definition, i.e., origin, destination, length, of blocks (groups of cars traveling together as a unit from the origin to the destination of the block; blocks result from classification operations at yards). Assets, engines, rail cars and, even, crews, assigned to a system based on full-asset-utilization operation policies can then "turn" continuously following circular routes and schedules (which include maintenance for vehicles and rest periods for crews) in the time-space service network, as illustrated in Figure 3 for a system with three yards and six time periods (Andersen *et al.*, 2006). The solid lines in Figure 3a represent services. There is one service from node 1 to node 3 (black arcs) and one service from node 3 to node 2 (grey arcs), both with daily frequency. Dotted arcs indicate repositioning moves (between different nodes) and holding arcs (between different time representations of same node). One feasible vehicle circuit in the time-space service network is illustrated in Figure 3b. The vehicle operates the service from node 3 to node 2, starting in time period 1 and arriving in time period 3. Then from period 3 to period 4 the vehicle is repositioned to node 1, where it is held for two time periods. In period 6 the vehicle operates the service from node 1 to node 3, arriving at time 1 where the same pattern of movements starts all over again.



3a. Service network 3b. Vehicle circuit

Figure 3. Full-asset-utilization-based service network and vehicle circuit

Freight carrier systems operating according to such policies require the same type of planning methods as when full-asset-utilization policies are not enforced. Yet, their particular characteristics lead to significant differences that require revisiting models, methods, and practices. The field is very new and, consequently, very little work has yet been done, particularly with respect to these new intermodal rail systems, which may be observed both in North America and Europe.

To illustrate differences and the corresponding challenges, consider that to adequately plan services according to a full-asset-utilization operating policy requires the asset circulation issue to be integrated into the service network design model. The requirement may be achieved by enforcing the condition that at each node of the network representation, i.e., at each yard and time period, the (integer) design flow must be balanced. Schematically, the service network design model of Section IV.2 becomes

$$\text{Minimize} \qquad \sum_{s \in S} F_s(y) + \sum_{p \in P} \sum_{l \in L_p} C_l^p(y,h) + \theta(y,h)$$

subject to $\sum_{l \in L_p} h_l^p = d_p \qquad p \in P,$

$$(y_s, x_l^p) \in \mathrm{X} \qquad s \in S,\ l \in L_p,\ p \in P,$$

$$\sum_{s \in S} y_{si^+} - \sum_{s \in S} y_{si^-} = 0 \qquad s \in S,\ i \in N,$$

$$y_s \geq 0 \text{ and integer} \qquad s \in S,$$

$$h_l^p \geq 0 \qquad l \in L_p,\ p \in P,$$

where si^+ and si^- indicate, respectively, that service $s \in S$ arrives and stops or terminates at, and that it initiates or stops and departs from node (yard) $i \in N$ in the appropriate period, while the third set of constraints enforces the balance of the total number of services arriving and departing at each yard and time period. Such requirements increase the difficulty of the planning problem. Pedersen, Crainic, and Madsen (2006) and Andersen *et. al.* (2006, 2007) present formulations and propose solution methods, but significant research work is still needed.

Many other issues have to be addressed and offer an exiting research perspective. Consider, for example, that although bookings tend to “smooth” out demand and decrease its variability, the stochasticity of the system is not altogether eliminated. Regular operations tend to be disrupted by a number of phenomena, including the fact that arrivals of ships in container port terminals are not regularly distributed and custom and security verification may significantly delay the release of containers. When this occurs, rail operations out of the corresponding port are severely strained: there might be several days without arrivals, followed by a large turnout of arriving containers. Optimization approaches (e.g., Crainic *at al*., 2006) may be used to adjust service over a medium-term horizon in such a way that a full-asset-utilization policy is still enforced, but a certain amount of flexibility is added to services to better fit service and demand. Such approaches may become even more effective when appropriate information sharing and container-release time mechanisms are implemented.

Acknowledgments

Partial funding for this project has been provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) through its Discovery Grants and Chairs and Faculty Support programs.

References

Andersen, J., Crainic, TG, and Christiansen, M., *Cycle-based Formulations for Capacitated Multicommodity Network Design with Asset Management Considerations*, Publication, CIRRELT, Montreal, Canada, 2007.

Andersen, J., Crainic, TG, and Christiansen, M., Optimizing Intermodal Transfers, *European Journal of Operational Research*, 2006 (forthcoming).

Andersen, J. and Christiansen, M., *Optimal Design of New Rail Freight Services*, Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, 2006.

Association of American Railroads, Class I Railroad Statistics, 2006, available on line at http://www.aar.org/PubCommon/Documents/AboutTheIndustry/Statistics.pdf

Bolis, S. and Maggi, R., Logistics Strategy and Transport Service Choices: An Adaptive Stated Preference Experiment, *Growth and Change*, 34(4): 490-504, 2003.

Bostel, N. and Dejax, P., Models and Algorithms for Container Allocation Problems on Trains in a Rapid Transshipment Shunting Yard, *Transportation Science* 32(4):370-379, 1998.

Campbell, J.F., Integer programming formulations of discrete hub location problems, *European Journal of Operational Research*, 72(2): 387-405, 1994.

Campbell, J.F., Ernst, and Krishnamoorthy, M., Hub Location Problems in Facility Location: Application and Theory, Drezner, Z. and Hamacher, H. Eds., Springer Verlag, Berlin, 373-407, 2002.

Christiansen, M., Fagerholt, K., and Ronen, D., Ship Routing and Scheduling: Status and Perspectives, *Transportation Science* 38(1):1-18, 2004.

Christiansen, M., Fagerholt, K., Nygreen, B., and Ronen, D., Maritime Transportation, Chapter 4 in *Transportation*, *Handbooks in Operations Research and Management Science*,. Barnhart and G. Laporte (Eds.), North-Holland, Amsterdam, 189-284, 2007.

Cheung, R.K. and Chen, C.-Y., A Two-Stage Stochastic Network Model and Solution Methods for the Dynamic Empty Container Allocation Problem, *Transportation Science* 32(2):142-162, 1998.

Cordeau, J.-F., Toth, P., and Vigo, D., A Survey of Optimization Models for Train Routing and Scheduling, *Transportation Science* 32(4): 380-404, 1998.

Crainic, T.G., Service network design in freight transportation, *European Journal of Operational Research*, 122(2): 272-288, 2000.

Crainic, T.G., Long-haul freight transportation, in *Handbook of Transportation Science*, Hall, R.W., Ed., Kluwer Academic Publishers, New York, 2nd Edition, 451-516, 2003.

Crainic, T.G. and Kim, K.H., Intermodal Transportation, Chapter 8 in *Transportation*, *Handbooks in Operations Research and Management Science*, C. Barnhart and G. Laporte (Eds.), North-Holland, Amsterdam, 467-537, 2007.

Crainic, T.G. and Laporte, G. Planning models for freight transportation, *European Journal of Operational Research*, 97(3): 409-438, 1997.

Crainic, T.G. and Rousseau, J.-M., Multicommodity, multimode freight transportation: A general modeling and algorithmic framework for the service network design problem, *Transportation Research Part B: Methodological*, 20(3): 225-242, 1986.

Crainic, T.G., Gendreau, M., and Dejax, P.J., Dynamic and Stochastic Models for the Allocation of Empty Containers, *Operations Research* 41(1):102-126, 1993.

Crainic, T.G., Bilegan, I.-C., and Gendreau, M., *Fleet Management for Advanced Intermodal Services - Final report*, Report for Transport Canada, Publication CRT-2006-13, Centre de recherche sur les transports, Université de Montréal, 2006.

Ebery, J, Krishnamoorthy, M., Ernst, A., and Boland, N., The Capacitated Multiple Allocation Hub Location Problem: Formulations and Algorithms, *European Journal of Operational Research* 120(3):614-631, 2000.

Harper, D.V. and Evers, P.T., Competitive issues in intermodal railroad-truck service, *Transportation Journal*, 32(3): 31-45, 1993.

European Conference of Ministers of Transport, *Terminology on Combined Transport*, page 44, United Nations, New York and Geneva, 2001, available online at http://www.cemt.org/online/glossaries/termcomb.pdf)

Evers, P.T. and Johnson, C.J., Performance perceptions, satisfaction and intention: The intermodal shipper's perspective, *Transportation Journal*, 40(2): 27-39, 2000.

Evers, P.T., Harper, D.V., and Needham, P.M., The determinants of shipper perceptions of modes, *Transportation Journal*, 36(2): 13-25, 1996.

Günther, H.-O. and Kim, K.H., Container Terminals and Automated Transport Systems, Springer-Verlag, Berlin, 2005.

Macharis, C. and Bontekoning, Y.M., Opportunities for OR in Intermodal Freight Transport Research: A Review, *European Journal of Operational Research* 153(2): 400-416, 2004.

McGinnis, M.A., The relative importance of cost and service in freight distribution choice: Before and after deregulation, *Transportation Journal*, 30(1): 12-19, 1990.

Murphy, P.R. and Hall, P.K., The relative importance of cost and service in freight distribution choice before and after deregulation: an update, *Transportation Journal*, 35(1): 30-38, 1995.

O'Kelly, M.E., A quadratic integer program for the location of interacting hub facilities, *European Journal of Operational Research*, 32(3): 393-404, 1987.

Pedersen, M.B., Crainic, T.G., and Madsen, O.B.G., *Models and Tabu Search Meta-heuristics for Service Network Design with Vehicle Balance Requirements*, Publication CRT-2006-22, Centre for Research on Transportation, Montréal, 2006.

Pedersen, M.B. and Crainic, T.G., *Optimization of Intermodal Freight Service Schedules on Train Canals*, Publication, Centre for Research on Transportation, Montréal, 2007.

Powell, W.B. and Carvalho, T.A., Real-Time Optimization of Containers and Flatcars for Intermodal Operations, *Transportation Science* 32(2): 110-126, 1998.

Powell, W.B. and Sheffi, Y., Design and Implementation of an Interactive Optimization System for the Network Design in the Motor Carrier Industry, *Operations Research* 37(1): 12-29, 1989.

Powell, W.B. and Topaloglu, H., Fleet Management, in *Applications of Stochastic Programming*, *Math Programming Society - SIAM Series on Optimization*, S. Wallace and W. Ziemba (Eds), SIAM, Philadelphia, PA, 185-216, 2005.

Powell, W.B., Jaillet, P., and Odoni, A., Stochastic and Dynamic Networks and Routing, in *Network Routing*, Volume 8 of *Handbooks in Operations Research and Management* Science, Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., Editors, North-Holland, Amsterdam, 141-295, 1995.

Powell, W.B., Bouzaïene-Ayari, B., and Simaõ, H.P. (2006), Dynamic Models for Freight Transportation, Chapter 5 in *Transportation*, *Handbooks in Operations Research and Management Science*, C. Barnhart and G. Laporte (Eds.), North-Holland, Amsterdam, 285-365, 2007.

Steenken, D., Voß, S., and Stahlbock, R., Container Terminal Operation and Operations Research - A Classification and Literature Review, *OR Spectrum* 26(1): 3-49, 2004.

Sussman, J. *Introduction to transportation systems*. Artech House. Boston, 2000.

Toth, P. and Vigo, D., *The Vehicle Routing Problem*, Volume 9 of *SIAM Monographs on Discrete Mathematics and Applications*, SIAM, 2002.

The Railway Association of Canada, *Railway Trends 2006*, Ontario, 2006, available on line at http://www.railcan.ca/documents/publications/1349/2006_10_24_RAC_Trends_en.pdf .

U.S. Department of Transportation, *Freight in America*, Washington D.C., 2006, available online at http://www.bts.gov/publications/freight_in_america/), page 24.