ELSEVIER

European Journal of Operational Research 97 (1997) 409–438

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Invited Review

Planning models for freight transportation

Teodor Gabriel Crainic [a,b], Gilbert Laporte [b,c,*]

[a] Département des sciences administratives, Université du Québec à Montréal, Canada

[b] Centre de recherche sur les transports, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Canada H3C 3J7

[c] Service de l'enseignement des méthodes quantitatives de gestion, École des Hautes Études Commerciales, Montréal, Canada

Received 1 September 1996

Abstract

The objective of this paper is to identify some of the main issues in freight transportation planning and operations, and to present appropriate Operations Research models and methods, as well as computer-based planning tools. The presentation is organized according to the three classical decision-making levels: strategic, tactic, operational. For each case, the problem and main issues are described, followed by a brief literature review and significant methodological and instrumental developments. We conclude with a few development perspectives. © 1997 Elsevier Science B.V.

Keywords: Freight transportation; Operations research; Planning; Operations; Models; Methods; Computer-based tools

1. Introduction

Transportation is an important domain of human activity. It supports and makes possible most other social and economic activities and exchanges. Freight transportation, in particular, is one of today's most important activities, not only as measured by the yardstick of its own share of a nation's gross national product (GNP), but also by the increasing influence that the transportation and distribution of goods have on the performance of virtually all other economic sectors.

A few figures illustrate these assertions (Cordeau and Laporte, 1996). It has been estimated (Taff, 1978) that transportation accounts for approximately 10% of the United States GNP and current figures could very well be significantly larger. In the United Kingdom, for example, transportation represents some 15% of national expenditures (Button, 1993). These figures are similar to those observed for Canada (some 16%; Zalatan, 1993) and France (around 9%; Quinet, 1990). Furthermore, transportation represents a significant part of the cost of a product. In Canada, for example, this part may reach 13% for the primary industrial sector and 11% for the transformation and production industry (Owoc and Sargious, 1992).

Transportation is also a complex domain, with several players and levels of decision, where investments are capital-intensive and usually require long implementation delays. Furthermore, freight transportation has to adapt to rapidly changing political, social and economic conditions and trends. It is thus a domain where accurate and efficient methods and tools are required to assist and enhance the analysis of planning and decision-making processes.

The objective of this article is to identify some of the main issues in freight transportation planning and operations and to present appropriate Operations Research models, methods, and computer-based planning

* Corresponding author.

PII S0377-2217(96)00298-6

tools. We start this presentation by recalling the contemporary environment of the freight transportation sector, its main players, as well as a general decision-making framework.

The freight transportation industry, as all other economic sectors, has to achieve high performance levels both in terms of economic efficiency and service quality. Economic efficiency because a transportation firm has to make profits while at the same time competing in an increasingly open and competitive market [1] where cost (or cost for a given quality level) is still the major decision factor in selecting a carrier or distribution firm. Yet, one also observes an increasing emphasis on the quality of the service offered. Indeed, the new paradigms of production and management, such as small or no inventories associated to just-in-time procurement, production and distribution, quality control of the entire logistics chain driven by customer demand and requirements, etc., impose high service standards on the transportation industry. This applies, in particular, to total delivery time (be there fast) and service reliability (be there within specified limits and be consistent in your performance).

The political evolution of the world also has an impact on the transportation sector. The emergence of free trade zones, in Europe and on the American continent in particular, has tremendous consequences for the evolution of freight transportation systems, not all of which are yet apparent or well understood. For example, open borders generally mean that firms are no longer under the obligation to maintain a major distribution center in each country. Then, distribution systems are reorganized and this often results in fewer warehouses and transportation over longer distances (which still have to perform according to low cost-high service standards). A significant increase in road traffic is a normal consequence of this process, as may be observed in Europe. A study conducted for the European Parliament forecasts a 34% increase in land-based transport for the countries of the European Economic Community between 1988 and the year 2000 (Button, 1993).

Additional factors which impact on the organization, operation policies and competitiveness conditions in the transportation industry are the internationalization of the economy and the opening of new markets due to political changes, mainly in central Europe and Asia, and the evolution of the regulatory environment. The first two imply larger economic spaces and transportation networks. Thus, from 1971 to 1988, the total volume of goods moved by ship has doubled, while the total number of kilometers covered by air cargo quadrupled (Button, 1993). Changes to the regulatory environment of transportation, particularly significant in North America and starting to gather momentum in Europe and elsewhere, also has a powerful impact on the operation and competitive environment of transportation firms. The deregulation drive of the 80's has seen governments remove numerous rules and restrictions, especially with regard to the entry of new firms in the market and the fixing of tariffs and routes, resulting in a more competitive industry and in changes in the number and characteristics of transportation firms. At the same time, more stringent safety regulations have been imposed, resulting in more complex planning and operating procedures.

There are several different types of players in the transportation field, each with its own set of objectives and means. Drawing a complete picture is well beyond the scope of this article. Here, we only identify some of the most important players. Producers of goods require transportation services to move raw materials and intermediate products and to distribute final goods in order to meet demands. Hence, they determine the demand for transportation and are often called *shippers*. (Other players, such as brokers, may also fall in this category.) Transportation is usually performed by *carriers*, such as railways, shipping lines, motor carriers, etc. Thus, one may describe an intermodal container service or a port facility as a carrier. Governments constitutes another important group of players. First, they still regulate several aspects of freight transportation (dangerous and obnoxious goods transportation, for example). They also provide a large part of the transportation infrastructure: roads and highways, and often a significant portion of the port, internal navigation, and rail facilities.

Transportation systems are rather complex organizations which involve a great deal of human and material resources and which display intricate relationships and trade-offs among the various decisions and management policies affecting their different components. It is then convenient to classify these policies

[1] Ash (1993) reports an average profit rate of only 2% for the Canadian motor carrier industry in 1992.

according to the following three *planning levels*:

(1) *Strategic* (long term) planning at the firm level typically involves the highest level of management and requires large capital investments over long time horizons. Strategic decisions determine general development policies and broadly shape the operating strategies of the system. Prime examples of decisions at this planning level are the design of the physical network and its evolution (upgrading or resizing), the location of main facilities (rail yards, multimodal platforms, etc.), resource acquisition (motive power units, rolling-stock, etc.), the definition of broad service and tariff policies, etc. Strategic planning also takes place at the international, national and regional levels, where the transportation networks or services of several carriers are simultaneously considered. State transportation departments, consultants, international shippers, etc. engage in this type of activity.

(2) *Tactical* (medium term) planning aims to ensure, over a medium term horizon, an efficient and rational allocation of existing resources in order to improve the performance of the whole system. At this level, data is aggregated, policies are somewhat abstracted and decisions are sensitive only to broad variations in data and system parameters (such as the seasonal changes in traffic demand) without incorporating the day-to-day information. Tactical decisions need to be made mainly concerning the design of the service network, i.e., route choice and type of service to operate, general operating rules for each terminal and work allocation among terminals, traffic routing using the available services and terminals, repositioning of resources (e.g., empty vehicles) for use in the next planning period.

(3) *Operational* (short term) planning is performed by local management (yardmasters and dispatchers, for example) in a highly dynamic environment where the time factor plays an important role and detailed representations of vehicles, facilities and activities are essential. Scheduling of services, maintenance activities, crews, etc., routing and dispatching of vehicles and crews, resource allocation are important operational decisions.

This classification highlights how the data flows among the decision-making levels and how policy guidelines are set. The strategic level sets the general policies and guidelines for the decisions taken at the tactical level, which determines goals, rules and limits for the operational decision level regulating the transportation system. The data flow follows the reverse route, each level of planning supplying information essential for the decision making process at a higher level. This hierarchical relationship prevents the formulation of a unique model for the planning of freight transportation systems and calls for different model formulations addressing specific problems at specific levels of decision making.

The remainder of the article is dedicated to presentation of important issues in planning and managing freight transportation systems. The presentation is organized according to the three planning levels just described. In each case, we identify the major players, describe the problem and the main related questions, review how the questions have been addressed through the development of operations research models, methods, and computer systems.

2. Strategic planning models

For a firm, strategic decisions determine general development policies and broadly shape the operating strategies of the system over relatively long time horizons. Several such decisions affect the design of the physical infrastructure network: where to locate facilities (loading and unloading terminals, consolidation centers, etc.), what type they should be (capacity, for example), on what lines to add capacity or which ones to abandon, etc. These issues, which may be collectively identified as *logistics system design*, are the subject of the first two sections where location and network design models are considered. In the third, the focus is broadened and we examine strategic planning issues and models aimed at the international, national and regional levels, where the transportation networks or services of several carriers are simultaneously considered.

2.1. Location models

Location problems involve the siting of one or several facilities, usually at vertices of a network, in order to facilitate the movement of goods or the provision of services along the network. The main location models are often classified as follows.

(1) *Covering models.* Locate facilities at the vertices of a network so that the remaining vertices are covered by a facility, i.e., they lie within a given distance of a facility. The problem can be to minimize the cost of locating facilities, subject to a constraint stating that all remaining vertices are covered. If one operates within a fixed budget, then an objective can be to maximize the demand covered by the facilities.

(2) *Center models.* Locate p facilities at vertices of a network in order to minimize the maximal distance between a vertex and a facility.

(3) *Median models.* Locate p facilities at vertices on the network and allocate demands to these facilities in order to minimize the total weighted distance between facilities and demand points. If facilities are uncapacitated and p is fixed, one obtains the so-called p-median problem. In such a case, each vertex is associated to its closest facility. If p is a variable and facilities are uncapacitated, this defines the *Uncapacitated Plant Location Problem* (UPLP). If p is a decision variable and facilities are capacitated, one obtains the *Capacitated Plant Location Problem* (CPLP).

Covering problems are typically associated with the location of public facilities such as health clinics, post offices, libraries, schools, etc. Center problems often arise in the location of emergency facilities such as fire or ambulance stations. Median problems are directly relevant to freight distribution. Here we will describe the CPLP. For a review of these and other models, see Daskin (1995) or Labbé, Peeters and Thisse (1995).

The CPLP can be formulated as follows. Assume there are n points in the plane called *vertices* and define

f_j: the cost of locating a facility at vertex v_j;
d_i: the demand at vertex v_i;
c_{ij}: the travel cost per unit of demand between vertices v_i and v_j;
u_j: the capacity of a facility located at vertex v_j;
y_j: a 0, 1 variable equal to 1 if and only if a facility is located at vertex v_j;
x_{ij}: the fraction of the demand of vertex v_i served by a facility located at a vertex v_j.

The model is then

$$\text{Minimize} \quad \sum_j f_j y_j + \sum_i \sum_j d_i c_{ij} x_{ij}, \tag{1}$$

$$\text{subject to} \quad x_{ij} \le y_j \quad \text{for all } i \text{ and } j, \tag{2}$$

$$\sum_j x_{ij} = 1 \quad \text{for all } i, \tag{3}$$

$$\sum_i d_i x_{ij} \le u_j y_j \quad \text{for all } j, \tag{4}$$

$$y_j = 0 \text{ or } 1 \quad \text{for all } i, \tag{5}$$

$$x_{ij} \ge 0 \quad \text{for all } i \text{ and } j. \tag{6}$$

In this model, the objective function represents the sum of fixed facility costs and transportation costs. It is assumed these costs are scaled over the same planning horizon. Constraints (2) express the condition that vertex v_i can only be served by vertex v_j if a facility is located at v_j. Constraints (3) state that the entire demand of each vertex must be allocated to facilities. Constraint (4) ensure that the capacity of a facility is never exceeded by its assigned demand.

The standard methodology for this model is to relax capacity constraints (4) in a Lagrangian fashion to obtain an objective of the form

$$\text{Minimize} \quad \sum_j \alpha_j y_j + \sum_i \sum_j \beta_{ij} x_{ij}, \tag{7}$$

where $\alpha_j = f_j - \mu_j u_j$, $\beta_{ij} = d_i c_{ij} + \mu_j d_i$, and μ_j are positive Lagrangian multipliers. The resulting problem is an Uncapacitated Plant Location Problem which can be solved by the DUALOC procedure devised by Erlenkotter (1978). For a detailed description of this approach, the reader is again referred to Daskin (1995).

More complex location problems arise in production and transportation planning. As an example, Crainic et al. (1989) present a multicommodity location–allocation with balancing requirements model. The formulation aims to determine the best logistics structure of the land distribution and transportation component of an international container shipping company, including the selection of inland depots, the assignment of customers to depots for each container type and direction of operations (allocation of empty containers from depots to customers and return of empties from customers to depots), and the determination of the main repositioning flows of empty containers to counter the regional differences between supplies and demands. The problem presents characteristics similar to a multicommodity Simple Plant Location Problem (the selection of the depots as discrete choice variables) with a complete multicommodity network flow structure (the allocation and repositioning of containers as continuous decision variables). Dual ascent methods, of DUALOC type, have been proposed (Delorme and Crainic, 1995), as well as branch-and-bound (Gendron and Crainic, 1995) and tabu search procedures (Crainic et al., 1993b).

2.2. Network design models

Network design problems are a generalization of location formulations. They are defined on graphs containing *nodes* or *vertices*, connected by *links*. Typically, links are directed and are represented by *arcs* in a network. When it is not necessary to specify a direction, they are represented by *edges*. Some of the vertices represent *origins* of some transportation demand for one or several commodities or products, while others (possibly the same) stand for the *destinations* of this traffic. Links may have various characteristics, such as length, capacity and cost. In particular, *fixed costs* may be associated to some or all links, signaling that as soon as one chooses to use that particular arc, one has to incur the fixed cost, in excess of the utilization cost which is in most cases related to the volume of traffic on the link. Note that when the fixed costs are associated to nodes, one obtains the formulations of a location problem. Such representations are generally used to model the cost of constructing new facilities, or of offering new transport services, or of adding capacity to existing facilities. In network design problems, the aim is to choose links in a network, along with capacities, eventually, in order to enable goods to flow between their origin and destination at the lowest possible system cost, i.e., the total fixed cost of selecting the links plus the total variables cost of using the network.

The simplest version of this problem is probably the *Shortest Spanning Tree Problem* (SSTP) which consists of determining a minimal length tree linking all vertices of an undirected graph $G = (V, E)$, where V is a vertex set and E is an edge set. To formulate the SSTP, define:

c_{ij}: the length of edge (v_i, v_j);

x_{ij}: a 0, 1 variable equal to 1 if and only if edge (v_i, v_j) belongs to the SSTP.

The problem is then:

$$\text{Minimize} \sum_{i<j} c_{ij} x_{ij}, \tag{8}$$

$$\text{subject to} \sum_{\substack{v_i \in S, v_j \in V \setminus S \\ \text{or} \\ v_i \in V \setminus S, v_j \in S}} x_{ij} \geq 1 \quad S \subset V,\ S \neq \emptyset, \tag{9}$$

$$x_{ij} = 0, 1 \quad \text{for all } (v_i, v_j) \in E. \tag{10}$$

Here, the objective function represents the total length of the tree, while constraints (9) state every subset S of nodes must be linked to its complement. Since it is uneconomical to over-connect the graph, there will be no cycles and the solution will be a tree.

The SSTP can be solved to optimality by applying one of several simple greedy algorithms (Ajuha et al., 1993). Here, we describe Kruskal's algorithm (Kruskal, 1956):

Step 1. Sort the edges in non-decreasing order of their lengths. Let T be the set of edges included in the tree. Set $T := \emptyset$. *Step 2.* Consider the next edge in the list. If it does not form a cycle with the edges of T, include it in T. If $|T| = |V| - 1$, stop. Otherwise, repeat this step.

More complex network design problems arise in transportation planning and operations (see also Section 4). In their most general formulation (Magnanti and Wong, 1984; Minoux, 1989; Magnanti and Wolsey, 1995; and the book by Ahuja et al., 1993), the models contain the same two types of variables defined in the previous formulations:

y_{ij}: 0, 1 variables modeling discrete choice design decisions; for a link (v_i, v_j), y_{ij} will equal to 1 if and only if link (v_i, v_j) is "opened";

x_{ij}^p: continuous flow decision variables indicating the amount of flow of commodity p using link (v_i, v_j).

The model then becomes:

$$\text{Minimize} \sum_{ij} f_{ji} y_{ji} + \sum_{ij} \sum_{p} c_{ij}^p x_{ij}^p, \tag{11}$$

$$\text{subject to} \sum_{j \in N} x_{ji}^p - \sum_{j \in N} x_{ij}^p = d_i^p \quad \text{for all } p \text{ and } i, \tag{12}$$

$$\sum_p x_{ij}^p \leq u_{ij} y_{ij} \quad \text{for all links } (v_i, v_j), \tag{13}$$

$$(y_{ij}, x_{ij}^p) \in S \quad \text{for all links } (v_i, v_j) \text{ and all } p, \tag{14}$$

$$y_{ij} = 0 \text{ or } 1 \quad \text{for all links } (v_i, v_j), \tag{15}$$

$$x_{ij}^p \geq 0 \quad \text{for all links } (v_i, v_j) \text{ and all } p, \tag{16}$$

where

f_{ij}: the fixed cost of "opening" link (v_i, v_j);

c_{ij}^p: the travel cost per unit of flow of product p on link (v_i, v_j);

w^p : the total demand of product p;

d_i^p: the demand at vertex v_i:

$$d_i^p = \begin{cases} -w^p & \text{if vertex } v_i \text{ is the origin of commodity } p, \\ w^p & \text{if } v_i \text{ is the destination of commodity } p, \\ 0 & \text{otherwise;} \end{cases} \tag{17}$$

u_{ij}: the capacity of link (v_i, v_j).

This is the linear cost, multicommodity version of the formulation. There exist important applications (e.g., variants of the service network design problem—see Section 3.1) that require nonlinear formulations, but this subject is beyond the scope of this paper. Similarly, we focus on multicommodity formulations since they represent the vast majority of applications in transportation and in other areas such as telecommunications and production.

In the network design formulation, the objective function (11) measures the total cost of the system. An interesting point of view is to consider this objective as also capturing the trade-offs between the costs of offering transportation infrastructure or services and those of operating the system. Eqs. (12), together with the demand definition (17), express the usual flow conservation and demand satisfaction restrictions. (Here, each commodity is associated to one origin-destination pair, but this assumption may easily be relaxed.) Constraints (13), often identified as *bundle* or *forcing* constraints, state that the total flow on link (v_i, v_j) cannot exceed its capacity u_{ij} if it is chosen in the design of the network, i.e., if $y_{ij} = 1$, and must be 0 if (v_i, v_j) is not part of the selected network, i.e., if $y_{ij} = 0$. Relations (15) and (16) specify the range of admissible values for each set of decision variables.

Eqs. (14) capture additional constraints related to the design of the network or relationships among the flow variables. Together, they may be used to model a wide variety of practical situations, and this is what makes network design problems so interesting. For example, the set S may represent topological restrictions imposed on the design of the network, such as precedence constraints (choose link (v_i, v_j) only if link (v_p, v_q) is chosen) or multiple choice constraints (select at most or exactly a given number of arcs out of a specified subset). An important type of additional constraint reflects the usually limited nature of available resources:

$$\sum_{(v_i, v_j)} f_{ji} y_{ji} \leq B. \tag{18}$$

These *budget* constraints illustrate a relatively general class of restrictions imposed upon resources shared by several (or all) links. Note that, quite often, budget constraints replace the fixed cost term in the objective function (Eq. (11)). *Partial capacity* constraints also belong to this group:

$$x_{ij}^p \leq b_{ij}^p \quad \text{for all links } (v_i, v_j) \text{ and all } p \tag{19}$$

and reflect restrictions imposed on the use of some facilities by individual commodities. Such conditions may be used to model, for example, the quantity of some hazardous goods moved by a train or a ship.

Much effort has been dedicated to uncapacitated versions of the problem and significant results have been obtained. The references already indicated provide an extensive review of the most significant results and real-life applications. In particular, Balakrishnan et al. (1989) present a dual-ascent procedure which, in conjunction with an add-drop heuristic, is capable of efficiently solving realistically sized instances of less-than-truckload consolidation problems (see Section 3.1). Less effort has been directed towards capacitated problems which are more difficult to solve and pose considerable algorithmic challenges. See Gendron and Crainic (1994, 1996) and Crainic et al. (1996) for a review of research in this area and for a presentation of bounding and tabu search procedures associated with this formulation.

2.3. Regional multimodal planning

Strategic planning activities must also to performed at a wider, regional, national or even international scale. The issues considered at this planning level usually concern the entire transportation system, or a significant part of it, the products that use it, as well as the interaction between passenger travel and freight flows.

The main questions addressed relate to the evolution of a given transportation system and its response to various modifications in its environment. Here we describe three such issues.

(1) What would be the impact on the system's performance of infrastructure modifications? Several types of modifications may affect the utilization and performance of a transportation system. One may build new facilities, or improve existing ones through modernization or capacity expansion, for example. One may also deconstruct the transportation network: the abandonment of underused, unprofitable rail lines is a typical example. These questions are often part of cost-benefit analyzes and of comparative studies of investment alternatives, especially when the available monetary resources are very limited, and are asked by regional or national planning authorities, as well as by international financial institutions such as the World Bank.

(2) How would the evolution of demand impact on the utilization of the system? Several changes may affect demand: (a) *Volume*: the quantities that have to be moved for each product, or product group, may increase or decrease in future years. (b) *Spatial distribution*: new communities appear and grow, new economic areas are developed, resources dry up in certain regions, the economic profile of a country or region evolves; these are only a few reasons why the spatial distribution of trade and freight flows may change in the future. (c) *Composition*: the relative importance of each product group in the trade exchanges between two zones varies according to the economic and social evolution (e.g., from rural to urban, from exporter of raw materials to a high technology industrial zone, etc.) of each of the two zones. Generally, one notices that several factors change simultaneously and one aims to evaluate the performance of the system under forecast demand conditions.

(3) What would be the impact of government or industry policies? Government policies may significantly affect the distribution of traffic and thus the utilization and performance of a transportation system. Common examples include energy pricing and taxation, mode imposition (in some energy importing countries legislation specifies that bulk goods have to be transported by energy efficient modes such as rail), infrastructure utilization pricing (such as the new trends in pricing highway utilization according to weight and distance), etc. Decisions made by firms also impact the performance of the transportation system: rail line abandonments which may change the ability of the transportation system to serve all regions of a country, mergers of carriers (the current trend of mergers and partnerships among carriers of various modes involved in container transportation is profoundly changing the rules by which intermodal transportation is operated), the introduction of new services (e.g., transcontinental double stack container trains), etc.

Planning and regulatory agencies at various levels of government are particularly interested in such issues (in the United States, in particular, the new intermodal legislation seriously challenges the state and regional transportation departments in this respect), as are international agencies involved in financing major projects in developing countries. Private firms are also interested in these questions, for example, companies involved in the financing of transportation infrastructures, or firms that plan and operate the distribution of goods using several transportation modes. In all cases, the focus is on the specific representation of several transportation modes, the corresponding intermodal transfer operations, the various criteria used to determine the movement of freight and the associated estimation of the traffic distribution over the transportation system considered.

The prediction of multicommodity freight flows over a multimode network is therefore an important component of transportation science and has attracted significant interest in recent years. One notes, however, that, perhaps due to the inherent difficulties and complexities of such problems, the study of freight flows at the national or regional level has not yet achieved full maturity, in contrast to passenger transportation where the prediction of car and transit flows over multimode networks has been studied extensively and several of the research results have been transferred to practice (see, for instance, Florian, 1984, 1986; Florian and Hearn, 1995; or the book by Sheffi, 1985).

A class of models well studied in the past for the prediction of interregional commodity flows is the *spatial price equilibrium model* and its variants. This class of models determines simultaneously the flows between *producing* and *consuming* regions, as well as the *selling* and *buying* prices. The transportation network is usually modeled in a simplistic way (bipartite networks) and these models rely to a large extent on the *supply* and *demand* functions of the producers and consumers, respectively. The calibration of these functions is essential to the application of these models and the transportation costs are unit costs or may be functions of the flow on the network. There have been so far few multicommodity applications of this class of models, with the majority of applications having been carried out in the agricultural and energy sectors in an international or interregional setting. See the review by Florian and Hearn (1995), or the book by Nagurney (1993).

The class of *network models* is generally considered to be more appropriate for the type of planning issues considered here. These formulations enable the prediction of multicommodity flows over a multimode network, where the physical network is modeled at a level of detail appropriate for a nation or a large region and represents the physical facilities with relatively little abstraction. The demand for transportation services is exogenous and may originate from an input-output model, if one is available, or from other sources, such as observed demand or scaling of observed past demand. The choice of mode, or subsets of modes, used is exogenous and intermodal shipments are permitted. In this sense, these models may be integrated with econometric demand models as well. The emphasis is on the network representation and on the proper representation of congestion effects in a static model to be used for comparative studies or for discrete time multiperiod analyses.

Several studies in the 1970's used rather simple network representations. Guélat et al. (1990) and Crainic et al. (1990b) review and discuss these efforts. Several studies were also aimed at extending spatial equilibrium models to include more refined network representations and to consider congestion effects and shipper-carrier interactions (see, for example, Friesz and Harker, 1985; Harker and Friesz, 1986a,b). This line of research has not, however, yet yielded practical planning models and tools. This is not the case of the model presented by Friesz et al. (1986). This is a sequential model which uses two network representations: an aggregate, user-perceived network which serves to determine, by using traffic equilibrium principles, the carriers chosen by the shippers, and detailed separate networks for each carrier, where commodities are transported at least total cost. This approach has proven quite successful to study the logistics of products (e.g., coal) where a limited number of "shippers" (e.g., the electric utilities in the United States and their suppliers in the exporting countries) strongly determine the behavior of the system.

The modeling framework we will present is based on the work of Guélat et al. (1990). The formulation does not consider shippers and carriers as distinct actors in the decisions made in shipping freight. The level of aggregation appropriate for strategic planning of freight flows results in origins and destinations that correspond to relatively large geographical areas and leads to the specification of supplies and demands representing, for each of the products considered, the total volumes generated by all the individual shippers. Furthermore, demand for strategic freight analyses are often determined from data sources (national freight flow statistics, economic input/output models) which enable the identification of the mode used, but do not contain information on individual shippers. It is thus assumed that shippers' behavior is reflected in the origin to destination product matrices and in the specification of the corresponding mode choice, as indicated in the following.

In this modeling framework a mode is a means of transportation having its own characteristics, such as vehicle type and capacity, as well as specific cost measures. Depending on the scope and level of detail of the contemplated strategic study, a mode may represent a carrier or part of its network representing a particular transportation service, an aggregation of several carrier networks, or specific transportation infrastructures such as highway networks or ports.

The base network is the network consisting of the nodes, links and modes that represent all possible physical movements on the available infrastructure. To capture the modal characteristics of transportation, a link a is defined as a triplet (i, j, m), where i is the origin node, j is the destination node, and m is the mode allowed on the arc. Parallel links are used to represent situations where more than one mode is available for transporting goods between two adjacent nodes. This network representation is compact and enables easy identification of the flow of goods by mode, as well as cost and delay functions by product and mode.

Furthermore, the network model resembles the physical network, since, for example, the rail and road infrastructures are physically different. Also, when on a physical link there are two different types of services, such as diesel and electric train services on rail lines, a separate link may be assigned to each service to capture the fact that they have different cost and delay functions.

Once the network representation is chosen, it is necessary to model intermodal shipments, and indeed to allow for mode transfers at certain nodes of the network and to compute the associate costs and delays. Intermodal transfers t at a node of the network are modeled as link to link, hence mode to mode, allowed movements and are represented as (i, m_1, j, m_2, k), where (i, m_1, j) stands for the incoming link, while (j, m_2, k) indicates the outgoing link of the transfer.

A path in this network consists of a sequence of directed links of a mode, a possible transfer to another mode, a sequence of directed links of the second mode, etc. Thus, a mode change is only possible at a transfer node. This representation also allows for the restriction of flows of certain commodities to subsets of modes (e.g., iron ore may be shipped only by rail and ship) to capture the mode restrictions that occur in the operation of freight networks and transshipment facilities.

A product is any commodity (collection of similar products), good or passenger, that generates a link flow. Each product p transported over the multimode network is shipped from some origins to some destinations of the network. The demand for each product for all origin-destination pairs is exogenous and is specified by a set of O/D matrices. The mode choice for each product is also exogenous and is indicated by defining for each of these O/D matrices a subset of modes allowed for transporting the corresponding demand. For example, one may indicate that the traffic out of certain regions has to use rail, while in other regions there is a choice between rail and barges. Let $g^{m(p)}$ be a demand matrix associated with product p, where $m(p)$ is the subset of modes that may be used to move this particular part of product p.

In the context of strategic planning of freight flows on a national or regional scale, the most efficient use of the transportation infrastructure is to carry the freight at the least total cost. Even though it is reasonable to assume that even in countries where a central authority controls and regulates the shipment of goods, a variety of circumstances in fact prevent the precise achievement of the goal of shipping at least cost, the model we present is based on the objective of minimizing total costs. The notion of cost is central to the model and is interpreted in the most general way, in the sense that it may have different components, such as monetary cost, delay (in terminals and on the lines of the network), energy consumption, noise and pollution level, risk (in case of incidents and accidents involving hazardous goods), etc.

The flows of product p on the multimode network are the decision variables of the model. Flows on links a are denoted by v_a^p and flows on transfers t are denoted by v_t^p. Cost functions, which depend on the volume of goods, are associated to the links and transfers of the network. For product p, the average cost functions $s_a^p(v)$, on links, and $s_t^p(v)$, on transfers, correspond to a given flow vector v. Then, the total cost of the flow on arc a for the product p is the product $s_a^p(v)v_a^p$; and the total cost of the flow on transfer t is $s_t^p(v)v_t^p$. The total cost of flows for all products over the multimode network is the function F that is to be minimized over the set of flows which satisfy the conservation of the flow and nonnegativity constraints:

$$F = \sum_{p \in P} \left(\sum_{a \in A} s_a^p(v)v_a^p + \sum_{t \in T} s_t^p(v)v_t^p \right). \tag{20}$$

In order to write these constraints, let $K_{od}^{m(p)}$ denote the set of paths that for product p lead from origin o to destination d using only modes in $m(p)$. The flow conservation equations are then

$$\sum_{k \in K_{od}^{m(p)}} h_k = g_{od}^{m(p)} \quad \text{for all } o, d, m(p), p, \tag{21}$$

where h_k is the flow on path k. These constraints specify that the total flow moved over all the paths that may be used to transport product p must be equal to the demand for that product The nonnegativity constraints are

$$h_k \geq 0, \quad \text{for all } k \in K_{od}^{m(p)}, o, d, m(p), p. \tag{22}$$

Eqs. (21) stand for the same type of constraints as Eqs. (12) in the network design formulations but are written in terms of path flow variables. The relation between arc flows and path flows is

$$v_a^p = \sum_{k \in K^p} \delta_{ak} h_k, \quad \text{for all } a, p, \tag{23}$$

where K^p is the set of all paths that may be used by product p, and

$$\delta_{ak} = \begin{cases} 1 & \text{if } a \in k, \\ 0 & \text{otherwise,} \end{cases}$$

is the indicator function that identifies the arcs of a particular path. Similarly, the flows on transfers are

$$v_t^p = \sum_{k \in K^p} \delta_{tk} h_k, \quad \text{for all } t, p, \tag{24}$$

where

$$\delta_{ak} = \begin{cases} 1 & \text{if } t \in k, \\ 0 & \text{otherwise.} \end{cases}$$

Transfer t belongs to path k if the two arcs that define the transfer belong to k. Then, the system optimal multiproduct, multimode assignment model consists of minimizing (20), subject to restrictions (21) to (22) with the definitional constraints (23) and (24). The algorithm developed for this problem exploits the natural decomposition by product and results in a Gauss-Seidel like procedure which allows the solution of large size problems in reasonable computational times (Guélat et al., 1990).

This network model allows for a detailed representation of the transportation infrastructure, facilities and services as well as the simultaneous assignment of multiple products on multiple modes. It captures the competition of products for the service capacity available, a feature of particular relevance when alternative scenarios of network capacity expansion are considered. On the other hand, the model is sufficiently flexible to represent the transport infrastructure of one carrier only.

The model, embedded in the STAN interactive-graphic system (Crainic et al., 1990a), is used by several agencies and organizations in a number of countries and has been applied successfully in practice. Crainic et al. (1990b) present the application of this methodology to the study of freight rail transportation, while several other applications are discussed in Guélat et al. (1990) and Crainic et al. (1990a), for example.

3. The tactical planning problem

When examining freight transportation, we distinguish between *producers* of goods who own or operate their own transportation fleet, and *carriers* who perform transportation services for various shippers. From a planning point of view, a more interesting classification differentiates between transportation operations that are mainly concerned with long distance movements of goods, such as rail transportation and less-than-truckload (LTL) trucking, and those that perform several pick up and delivery operations, mainly by truck, over relatively short distances. The first case is often referred to as the *service network design* problem, while the second type of operations are usually identified as *vehicle routing problems*.

In all cases, we are considering transportation systems where one vehicle (e.g., truck, railway wagon, ship, etc.) or convoy (e.g. rail, truck or barge "trains") may serve to move freight of different customers with possibly different initial origins and final destinations. *Consolidation*-type operations are thus central to such systems and one of the main issues in tactical model development. This is in contrast to the "door-to-door" transportation operations performed, for example, by truckload motor carriers or by intermodal container transportation firms, and addressed in the next section.

When demands of several customers are served simultaneously by using the same "vehicle", one cannot tailor a service for each customer individually. Carriers have to establish regular services (e.g., a container ship from Seattle to Singapore) and adjust their characteristics (route, intermediary stops, frequency, vehicle and convoy type, capacity, speed, etc.) to satisfy the expectations of the largest number of customers possible. Externally, the carrier then proposes a series of services (or *routes* grouped into a *network*), each with its operational characteristics. Internally, the carrier builds a series of rules and policies that affect the whole system and are often collected in an *operational* (or *load* or *transportation*) plan. The aim is to ensure that the proposed services are performed as stated (or as close as possible) while operating in a rational and efficient way.

Customers' expectations have traditionally been expressed in terms of "going there" at the lowest cost possible. This, combined with the usual cost consciousness of any firm, has implied that the primary objective of a freight carrier was, and still is for many carriers, to operate at the lowest possible cost. Increasingly, however, customers not only expect low tariffs, but also require a high quality service, mostly in terms of speed, flexibility and reliability. The significant increase in the market share achieved by motor carriers, mainly at the expense of railway transportation, is due to a large extent to this phenomenon. Consequently, tactical planning not only aims at an adequate allocation and utilization of existing resources, but also strives to achieve the best trade-off between operating costs and service performance.

Tactical planning thus appears as a vital link in the planning process of a freight transportation carrier. Its output, the *transportation plan*, is used to determine the day-to-day policies that guide the operations of the system and is also a privileged evaluation tool for "what-if" questions raised during strategic planning. In the following, we examine models and methods that may be used to build the transportation plan in each of the two cases identified at the beginning of the section.

3.1. Service network design for intermodal transportation

In intercity freight transportation, service design is particularly relevant to firms and organizations which both supply or regulate transportation services and control, at least partially, the routing of goods through the service network. The presence of terminals where cargo and vehicles are consolidated, grouped or simply moved from one service to another further characterizes this type of transportation. Main examples of such transportation systems are:

(1) Transportation by rail where various train services (e.g., normal, rapid, direct, unit, etc.) correspond to various "modes".

(2) Less-than-truckload trucking, eventually incorporating multi-trailer assemblies and use of rail transportation.

(3) Container transportation through a combination of air, sea, road and rail modes.

(4) Freight transportation in developing countries where a central authority more or less controls a large part of the transportation system.

The underlying structure of any large freight transportation system consists of a rather complex network of terminals connected by physical links (e.g., rail tracks) or conceptual links (e.g., sea or truck lines). On this network, freight demand is specified by commodity class according to its origin and destination, in addition to physical and service characteristics, and is moved by carrier services performed by a large number of vehicles (e.g., railcars, trailers, etc.). Vehicles move, usually on specified routes and sometimes following a given schedule, either individually or grouped in convoys such as trains or multi-trailer assemblies. Convoys are formed and dismantled in terminals. Also in terminals, freight may be consolidated, loaded in and unloaded from vehicles, vehicles may be changed from one convoy to another, etc. Consequently, terminals come in several types and sizes. For railways, for example, one identifies large and small classification *yards* where railcars are consolidated into blocks and trains are formed, pick-up and delivery stations, junction points, etc. Similarly, an LTL network may encompass *end-of-line* terminals where the local traffic is delivered (by smaller pick-up trucks—see Section 3.2) and consolidated into larger shipments while loads from other parts of the network are unloaded and moved into smaller delivery trucks (see Section 3.2), and *breakbulk* terminals where traffic from many end-of-line terminals is unloaded, sorted and consolidated for the next portion of the journey. (Rail yards and breakbulks may further be classified according to their importance and role: regional, national, specialized services, etc.) The traffic on the links of an intermodal transportation network thus represents vehicle and, eventually, convoy movements, while operations are performed at terminal nodes on traffic, vehicles and convoys.

To further clarify these notions, consider the case of railway transportation. Here, everything begins when an order for a number of empty vehicles is issued by a customer or, alternatively, when freight is brought into the loading facility following a pick up operation. At the appropriate yard, railcars are selected, inspected and then delivered to the loading point. Once loaded, cars are moved to the origin yard (possibly the same) where they are sorted, or *classified*, and assembled into *blocks*. A block is a group of cars, with possibly different final destinations, arbitrarily considered as a single unit for handling purposes from the yard where it is made up, to its destination yard where its component cars are to be resorted. Rail companies use blocks as a mean to take advantage of some of the economies of scale related to full train loads and to the handling of longer car strings in yards.

The block is eventually put on a *train* and this signals the beginning of the journey. During the long haul part of this journey, the train may overtake other trains or be overtaken by trains with different speeds and priorities. When the train travels on single-track lines, it may also meet trains traveling in the opposite direction. Then, the train with the lowest priority has to give way and wait on a side line, for the train with the higher priority to pass by. At the yards where the train stops, cars and engines are regularly inspected. Also, blocks of cars may be *transferred*, i.e., taken off one train and put into another. When a block finally arrives at destination, it is taken off the train, its cars are sorted and those having reached their final destination are directed to the unloading station. Once empty, the cars are prepared for a new assignment, which may be either a loaded trip or an empty movement.

One source of complication in rail freight transportation is the complex nature of yard activities, in particular its main operations: the classification of cars and the composition of trains. The modeling of yard operations and of their interactions with the remainder of the system is a critical component of any comprehensive rail model. It is interesting to note that, traditionally, in most rail systems cars spend most of their lifetime in yards: being loaded and unloaded, being classified, waiting for an operation to be performed or for a train to come or, simply, sitting idle on a side track. Also of interest is the fact that most rail companies have dedicated yards to intermodal services in an attempt to cut on the delays associated with yard operations.

Less-than-truckload motor carrier transportation follows the same basic operational structure but on a simpler scale and with significantly more flexibility due to the fundamental difference in infrastructure: while rail transportation is "captive" of the rail tracks, trucks may use any of the existing links of the road and highway network as long as they comply with the weight regulations. Furthermore, a truck is only formed of a tractor and one or several trailers (when more than one trailer is used, these are smaller and are called "pups"). Consequently, the terminal operations are also significantly simpler: only freight is handled to consolidate outbound movements. Transfer operations may still take place, however, since trailers may be exchanged between convoys. Actually, this operation is performed quite regularly since not only does it not require any specialized equipment (most switches take place at truck stops along highways) but it also facilitates the building of efficient driver schedules.

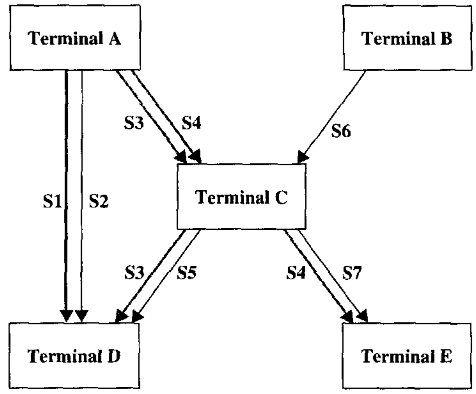

Fig. 1. Service network example.

To further clarify the preceding description and to illustrate the type of decisions and trade-offs characteristic of tactical planning, we refer to Fig. 1 which displays in a very simplified form part of an intermodal service network (rail or LTL, for example). Five terminals make up this network and seven *services* can be offered: one express direct service (S1), four normal direct services (S2, S5, S6 and S7) and two services with intermediary stops (S3 from terminal A to D with a stop at terminal C and S4 from A to E with a stop at C). Not indicated on the figure is the physical route followed by each service. Thus, it may happen that services S1, S2 and S3 follow exactly the same route, their differences being only at the level of their respective speeds (and, eventually other characteristics such as capacity, preferred traffic and priority) and intermediary stops. Let us examine the transportation options for a shipment of a given commodity from terminal A to terminal D. The shipment is classified at terminal A and it may be

(1) consolidated with other shipments going directly to D and put on a direct service, S1 or S2, according to its characteristics;

(2) consolidated into a load for D but put into a vehicle on service S3 which stops at terminal C to drop and pick up traffic;

(3) consolidated into a load for terminal C, put on service S4, unloaded at terminal C, reclassified and consolidated together with traffic originating in B

and C into a load for terminal D which will be put on service S5.

Which alternative is the "best"? Each has its own cost and delay characteristics which are a direct consequence of the service characteristics of each terminal and service. Thus, if the direct services from A to D are offered only rarely (due to generally low level of traffic demand, for example), it may be more efficient to consolidate the A to D traffic together with the A to C and A to D volumes and ship through terminal C. This would probably result in a higher utilization of the equipment and in a decrease of the waiting time at the original terminal, hence in a more rapid service for the customer. However, the same decision would also result in additional unloading, consolidation and loading operations, creating additional delays and a higher congestion level at terminal C, as well as a decrease of the total reliability of the shipment. On the other hand, to increase the frequency of a direct service from A to D would imply a faster and more reliable service for the corresponding traffic, as well as a decrease in the level of congestion at terminal C, but would also require additional resources which increases direct costs. Thus, to select the " best" solution for the customer and the company, one has to simultaneously consider the routing of all traffic demands, as well as the costs and service characteristics of each terminal and line operations.

More formally, the main decisions made at the tactical level concern the following issues:

(1) *Service network design*. Selection of the routes (origin and destination terminal, physical route and intermediate stops) on which services will be offered and the determination of the characteristics of each service, particularly their *frequency*.

(2) *Traffic distribution*. Routing specification for the traffic of each origin-destination pair (usually according to the specific commodity): services used, terminals passed through and operations performed the terminals.

(3) *Terminal policies*. General rules specifying for each terminal the type of consolidation to perform (for a rail application, these rules would specify the blocks into which cars should be classified, the trains that are to be formed and the blocks that should be put on each train, etc.). An efficient allocation of work among terminals is an important policy objective.

(4) *Empty balancing*. How to reposition empty vehicles to meet the forecast needs of the next planning period.

(5) *Crew and motive power scheduling*. How to allocate and reposition in preparation for the next planning period the resources required by the selected transportation plan.

These problems and decisions have network-wide impacts and are strongly and complexly interconnected both in their economic aspects and their space-time dimensions. Thus, the strategies developed for different planning problems interact and mutually influence each other. Furthermore, trade-offs have to be made between operating costs resulting from a given policy and service quality as measured, in most cases, by delays incurred by freight and rolling-stock, or by the respect of predefined performance targets. These trade-offs are important when a single problem and decision is considered and even more so when the relationships among decisions for different problems are contemplated. Therefore decisions should be made globally, network-wide, in an integrated manner (Crainic and Roy, 1988).

Several efforts have been directed toward the formulation of tactical planning models. Network models, that take advantage of the structure of the system and integrate policies affecting several terminal and line operations, are the most widely developed. These formulations may be classified into two main groups: network simulation and optimization models.

Rail companies have used simulation models for quite a long time (see Assad 1980, and the references quoted in this work). Models of this type simulate the movements of trains and cars through the rail network, given a set of operating policies for the yards and lines of the system and a set of train schedules. The user has to input, among other data, the classification and train formation policies at each yard, a complete set of train schedules and traffic demand information. The results are detailed cost information, an evaluation of the occupancy of each facility in the network (yards and lines) and sometimes an estimation of transit traffic times. Based on these results, the user may evaluate a given operating policy in detail. The main limitation of this approach is that simulation models are not able to generate new operating strategies that would incorporate a network-wide analysis of their impact and the evaluation of a number of apparently conflicting objectives. Also, simulation models usually require prohibitive data input and running times, which makes their repeated use with different sets of operating policies impractical.

Network optimization models, on the other hand, are less detailed but offer the advantage of fast generation, evaluation and selection of integrated, network-wide operating strategies with respect to some objective function, usually involving both operating costs and service criteria. See Crainic (1988) for a comprehensive review of network, yard and line models in rail tactical planning, and Delorme et al. (1988) for a similar review of models and methods for LTL transportation. Interesting recent contributions to the field are found in Farvolden and Powell (1994), Keaton (1989, 1991, 1992), Haghani (1989), Jovanovic and Harker (1991), Powell and Sheffi (1989), Braklow et al. (1992), Roy and Delorme (1989), Barnhart (1995).

The network optimization model that follows is based on the work of Crainic (1982) and Crainic et al. (1984). It integrates the service network design and traffic multimode routing problems with general terminal policies. Its goal is the generation of global strategies to improve the cost and service performance of the system. It is a modeling framework for the tactical planning problem of intermodal freight transportation systems in the sense that while it may represent a large variety of real situations, it has to be adapted to each application. In fact, it has been applied to problems from the Canadian (Crainic, 1984) and French (Crainic and Nicolle, 1986) railways. It has also been adapted with considerable success to the multimode LTL trucking problem (Roy, 1984; Delorme and Roy, 1989). In the following, we only present a simplified model in order to emphasize the main issues and challenges of formulating tactical planning models and tools.

Let $G = (V, E)$ represent the "physical network". Vertices of V represent the terminals selected for the particular application (terminals A, B, C, D and E of Fig. 1). For simplicity, assume that all terminals can perform all operations. The set E is the set of links representing the connections between terminals. The *transportation demand* is defined in terms of volume (e.g., number of vehicles) of a certain commodity c to be moved from an origin node $o \in V$, to a destination $d \in V$. To simplify, we refer to the market or *traffic-class* $m = (o, d, c)$ with a positive transportation demand d^m.

The *service network* specifies the transportation services that could be offered to satisfy this demand. Each *service* s in this structure is defined by the route it follows through the physical network from its origin to its destination, by the sequence of intermediate terminals, and by its service characteristics: mode, speed, capacity, etc. For example, in Fig. 1 service S3 may represent a normal rail train service from terminal A to terminal D with an intermediate stop at terminal C. Similarly, for an LTL application, service S2 could represent the operation of a direct truck between terminals A and D, while service S1 could stand for the option to move the trailer TOFC, taking advantage of existing long distance railway connections. The *frequency* F_s is the other important service characteristic. It defines the level of service offered on the route: how often is the service run during the planning period. A tactical model determines the (integer) values of the frequencies of the selected services, sometimes subject to lower (more rarely, upper) limits. Note that Fig. 1 actually illustrates a service network since the physical routes are not displayed.

Traffic moves following predefined *itineraries* that specify the paths of services to use and the sequence of intermediate terminals on this path where operations are to be performed. For a given traffic-class, one or several such itineraries may be used, according to the level of congestion in the system and the service and cost criteria of the particular application. The *routing* of freight (also called the distribution of freight), as given by the amount of flow X_k^m of traffic-class m moved by using itinerary k, must also be determined by the model.

Frequencies and itineraries are the central elements of the model. Fixing frequencies values determines the design of the service network, the level of service and the feasible domain for the traffic distribution problem. On the other hand, the selection of the best itineraries for each traffic class solves the traffic distribution problem and also determines the workloads and the general consolidation strategies for each terminal of the system.

The model states that the total system cost has to be minimized, while satisfying the demand for transportation and the service standards. It also contains the usual nonnegativity and integrality constraints:

$$\text{Minimize} \quad \Psi(X_k^m, F_s),$$
$$\text{subject to} \quad \sum_k X_k^m = d^m \quad \text{for all } m,$$
$$X_k^m \geq 0 \quad \text{for all } k, m,$$
$$F_s \geq 0 \quad \text{and integer for all } s.$$

The objective function $\Psi(X_k^m, F_s)$ defines the total system cost and includes (1) the total cost of operating a given service network at level F_s; this is the "fixed" cost of the transportation system; and (2) the total cost of moving freight by using the selected itineraries for each traffic-class (X_k^m); this is the "variable" cost of satisfying demand by a service network operated at level F_s.

What the objective function represents is a generalized cost, in the sense that both operating and service costs are included. It is at this level that the relationships and trade-offs among the various system and policy components are considered. Indeed, the objective combines the real costs of handling and moving freight and vehicles, and penalties related to service reliability (generally based on the mean and variance of transportation delays). These delays are incurred by vehicles, convoys and freight due to congestion and operational policies in terminals and on the road. The resulting model has the structure of a nonlinear, mixed integer, multimode, multicommodity flow problem, and may be solved by an efficient heuristic algorithm (Crainic and Rousseau, 1986) based on decomposition, column generation and descent techniques.

Eq. (25) illustrates one approach for integrating service considerations into the total generalized system cost:

$$\begin{aligned}\Psi(X_k^m, F_s) = &\sum_s C_s^t F_s + \sum_{m,k} C_{mk}^t X_k^m \\ &+ \sum_s C_s^d E(T_s) F_s \\ &+ \sum_{m,k} C_{mk}^d E(T_k^m) X_k^m \\ &+ \sum_s C_s^p (\min\{0, \alpha_s F_s - X_s\})^2. \end{aligned} \tag{25}$$

Here, delays are converted into "delay costs", compatible with operating costs (C_s^t and C_{mk}^t for each service and itinerary, respectively), via user-defined unit time costs for each traffic-class (C_{mk}^d) and type of service (C_s^d). These cost are usually based on equipment depreciation values, goods inventory costs and time-related characteristics, such as priorities or different degrees of time sensitivity for specific traffic classes. Eq. (25) also illustrates the use of penalty terms to capture various restrictions and conditions. Thus the service capacity restrictions are considered as utilization targets and over-assignment of traffic is permitted at the expense of additional costs and delays. Then, while solving the resulting mathematical programming problem, trade-offs between the cost of increasing the level of service and the extra costs of insufficient capacity may be addressed.

The terms most likely to appear in such a function, and especially in the computation of the mean delays $E(T_s)$ and $E(T_k^m)$ (for services and itineraries, respectively), are application specific. Generally, one attempts to include those that reflect the cost and delay characteristics of the terminal and line operations most significant for the system. Typically, for a rail application, one may have:

(1) Handling costs associated with car classification operations at yards (sorting, blocking, etc.).

(2) The cost of transferring cars or blocks among trains at terminals.

(3) The costs of breaking down and of making up trains at yard terminals.

(4) Hauling cost (by service class) for trains and cars over the lines of the network; these costs may relate to energy, motive power and crews.

(5) The average (and possibly the variance) of the delay due to yard operations: car and train inspection, car classification and blocking, connection delays (the waiting time for the designated service to be available), train formation, etc.

(6) The mean delays incurred by trains on the lines of the network due to congestion conditions and meet overtake operations. Stopping times at intermediate yards may also be included.

The same main elements are found in LTL applications, with the normal adjustment for the specificity of trucking operations:

(1) Costs and durations related to loading and unloading freight at terminals.

(2) Costs and delays due to transdock operations and consolidation functions at terminals.

(3) Waiting for departure and, sometimes, for an unloading gate to become available.

(4) Cost and time required to move vehicles and cargo from one terminal to another (note that congestion might appear even in this context when terminals are located within the congested part of urban highway and road networks).

(5) The associated manpower costs in terminals and for driving the vehicles.

Queueing models are generally used to derive classification, consolidation, connection and other terminal delay functions as well as over-the-road delays which reflect the congestion and physical characteristics of the system. For a review of such models, the interested reader may consult Crainic (1988), as well as the more recent contributions of Chen and Harker (1990), Harker and Hong (1990) and references quoted by these authors.

A second approach to integrating service considerations into the total generalized system cost is illustrated by Eq. (26) (Roy, 1984):

$$\begin{aligned}\Psi(X_k^m, F_s) = &\\ &\sum_s C_s^t F_s + \sum_{m,k} C_{mk}^t X_k^m \\ &+ \sum_{m,k} C_{mk}^S (\min\{0, S_m - E(T_k^m) - n\sigma(T_k^m)\}) X_k^m \\ &+ \sum_s C_s^p (\min\{0, \alpha_s F_s - X_s\})^2. \end{aligned} \tag{26}$$

In this case, the operating cost coefficients (C_s^t and C_{mk}^t) for vehicle (service) and freight handling and transport have the same general meaning as previously, as does the second penalty term of the equation. The first penalty term, appearing on the second line of Eq. (26), represents the case when service quality targets are announced. Here, for example, each traffic-class has an associated delivery objective (24 hours for 90% of deliveries, for example) and a penalty for not achieving it. Then, the service target may be compared with the mean expected delay of the entire journey. Trade-offs may be achieved during optimization between the costs of improving operations and those of achieving the promised service quality. Furthermore, scenario analyzes may be conducted following the optimization phase to examine, for example, the effect of the relative importance of each factor on the global performance of the system. For further examples of such postoptimal analyzes see Crainic and Roy (1988) and Roy and Crainic (1992).

An additional planning issue particularly important and challenging for freight carriers is the need to move empty vehicles. Indeed, the geographic differences in demands and supplies for each commodity type often result in an accumulation of empty vehicles in a region where they are not needed, or in a deficit of vehicles in other regions that require them. Freight carriers must therefore reposition vehicles for use in the following planning periods. This is a complicated issue and numerous studies reflect the significant research and development effort that has been dedicated to it. *Empty balancing* models are often associated with the operational level of planning. Interested readers may start exploring this field with the review of Dejax and Crainic (1987). This issue must also be considered at the tactical level. In rail transportation, for example, empty rail cars are put on the same trains as loaded ones and thus contribute to an increase in the number of trains, in the volume of vehicles handled in terminals and, ultimately, in system costs and delays. For planning purposes, the demand for empty cars may be approximated and introduced in the tactical model by viewing empties as another commodity to be transported. The issue is also relevant in LTL trucking where empty balancing is an integral part of a transportation plan. In this case, a transportation plan is first obtained for the actual traffic demands, and an empty balancing model is then solved to reposition the empties (see, e.g., Delorme and Roy, 1989; Braklow et al., 1992).

3.2. *Vehicle routing problems*

The models just described typically apply to long-distance freight transportation, where sorting and consolidation occurs at freight terminals. The next planning level takes place in more restricted geographical areas. It involves the distribution of goods at the local or the regional level and comprises activities such as pick-up, delivery, or a combination of both. In some industries, such as in the food and drink business, distribution costs at this level can account for up to 70% of the value added costs of goods (Golden and Wasil, 1987).

Distribution management problems arising at this level have been extensively studied by operations researchers over the last forty years. Interesting references are those of Eilon et al. (1971), of Bodin et al. (1983), of Golden and Assad (1988) and of Daganzo (1971). Such problems are generally referred to as *Vehicle Routing Problems* (VRPs), but this designation covers a wide range of setups rather than a specific problem. VRPs involve the design of pick-up or delivery routes from one or more central depots to a set of geographically scattered customers. Several versions of the problem can be defined, depending on a number of factors, constraints and objectives.

(1) Does the problem involve deliveries, collections, or a combination of both? Are there precedence relations between deliveries or collections?

(2) Does distribution take place from a single depot or from several centers?

(3) How many vehicles are involved? Is the number fixed or does it constitute a decision variable? Is the vehicle fleet homogeneous or heterogeneous? What are the capacity, speed, operating costs of these vehicles?

(4) What are the work conditions of the drivers? What is the pay structure? What is the length of a normal workday? What are the conditions on overtime? Are multiple same day trips allowed?

(5) Is the demand known in advance or is it revealed in a dynamic fashion during the course of operations?

(6) How often or when must each customer be visited during the planning period (one week, say)? On a given day, must customers be visited within specific time windows?

These are just some of the questions that must be addressed when solving a VRP. For a broader coverage of these issues, see Assad (1988). It is therefore important to have a clear understanding of the situation at hand and of the rules of the game.

How does one solve vehicle routing problems? Unfortunately, even for the most basic version of the problem, exact methods can handle only relatively small instances. To this day, no optimization algorithm can *consistently* solve instances involving more than 50 customers. In practice (and in all known VRP software), the only option is to use heuristics. There exists an abundant literature on exact and approximate algorithms for the VRP. For recent surveys, see Laporte (1992b), Fisher (1995), as well as Desrosiers et al. (1995).

To gain more insight into the nature of the problem and into some of the better known algorithms, we will focus on a very basic, but commonly studied, version of the problem: the single depot capacitated VRP (or CVRP). The CVRP is usually defined on an undirected graph $G = (V, E)$, where $V = \{v_0, v_1, \ldots, v_n\}$ is the vertex set, and $E = \{(v_i, v_j) \mid v_i, v_j \in E, i < j\}$ is the edge set. Vertex v_0 represents a depot at which are based m identical vehicles of capacity Q, while the remaining vertices represent customers. The demand of customer v_i is equal to a positive integer q_i. With each edge (v_i, v_j) is associated a cost c_{ij}, typically proportional to distance or travel time. The CVRP consists of designing m vehicle routes starting and ending at the depot, in such a way that each customer is visited exactly once by only one vehicle, the demand on any vehicle route does not exceed Q and the total cost is minimized. Several formulations have been proposed for this problem (for a survey, see Laporte and Nobert, 1987). Here we describe a simple and commonly employed "two-index vehicle flow formulation". Let x_{ij} be an integer variable equal to 2 if a vehicle makes a return trip between v_i and v_j, equal to 1 if it makes a single trip between v_i and v_j, and equal to 0 otherwise. The case x_{ij} can only occur if one of the two vertices is the depot. In what follows, x_{ij} must be interpreted as x_{ji} whenever $i > j$. The CVRP is then formulated as follows.

$$\text{Minimize} \sum_{v_i, v_j \in V} c_{ij} x_{ij}, \tag{27}$$

$$\text{subject to} \sum_{j=1}^{n} x_{0j} = 2m, \tag{28}$$

$$\sum_{i<j} x_{ik} + \sum_{j>k} x_{kj} = 2 \quad v_k \in V \setminus \{v_0\}, \tag{29}$$

$$\sum_{\substack{v_i \in S, v_j \in V\setminus S \\ \text{or} \\ v_j \in S, v_i \in V \setminus S}} x_{ij} \geq 2 \left\lceil \sum_{v_i \in S} q_i / Q \right\rceil$$

$$S \subseteq V \setminus \{v_0\},\ 2 \leq |V| \leq n-2, \tag{30}$$

$$x_{0j} = 0, 1 \text{ or } 2 \quad v_j \in V \setminus \{v_0\}, \tag{31}$$

$$x_{ij} = 0 \text{ or } 1 \quad v_i, v_j \in V \setminus \{v_0\}. \tag{32}$$

In this formulation, the objective function states that the total travel cost must be minimized. Constraints (28) express the requirement that each of the m vehicle trips must start and end at the depot, while constraints (29) state that the degree of every customer vertex must be equal to 2. Constraints (30) are connectivity constraints. They can be interpreted as follows. Given a set $S \subseteq V \setminus \{v_0\}$ of customers and their total demand $\sum_{v_i \in S} q_i$, in any feasible solution there must be at least $\lceil \sum_{v_i \in S} q_i/Q \rceil$ vehicles servicing S, and twice that number connecting S to its complement $V \setminus S$. These constraints play a double role: they ensure that all customers are connected to the depot (since the right-hand side is never less than 2), and also that the total demand of any route never exceeds the vehicle capacity. In constraints (31), the case $x_{0j} = 2$ corresponds to a return trip between the depot and customer v_j. In all other cases, x_{ij} is equal to 0 or 1. One interest of this formulation is that m, the number of vehicles, can be regarded as a constant or as a variable. In the latter case, one can add to the objective function an extra term fm, where f is the fixed cost of a vehicle.

The main drawback of this formulation is that the number of integer variables is relatively large in most practical situations and the number of connectivity constraints is exponential in n. Therefore this model is never solved directly. It is usually tackled by means of a *branch-and-cut* algorithm. Initially, a simplified version of the problem without integrality requirements and without connectivity constraints is solved. A heuristic is applied to determine whether any of the connectivity constraints are violated. If this is the case, several of the most violated constraints are introduced into the model which is then reoptimized. If no violated connectivity constraint can be identified, a check for integrality is made. If the solution is feasible, it is then optimal. Otherwise, two subproblems are created by branching on a fractional variable x_{ij}. A search tree is thus created and the same process is reapplied at each node of the search tree, by following the usual branch-and-bound rules. This method was first applied by Laporte et al. (1985) who solved to optimality some loosely constrained 60- customer problems (with vehicles filled at about 70% of their capacity). The method does not perform so well on tightly constrained problems because of the very large number of violated connectivity constraints that have to be generated. Recently, a number of other valid, but more complicated, constraints have been identified for the CVRP. Some of these results are presented in Cornuéjols and Harche (1993).

Countless heuristics have been proposed for the CVRP. We will first sketch two classical methods: the savings method (Clarke and Wright, 1964) and the sweep heuristic (Gillett and Miller, 1974). These two algorithms are based on intuitive principles and have gained a fair degree of acceptance over the years. We will then briefly describe a more recent and more powerful algorithm based on tabu search (Glover, 1989, 1990).

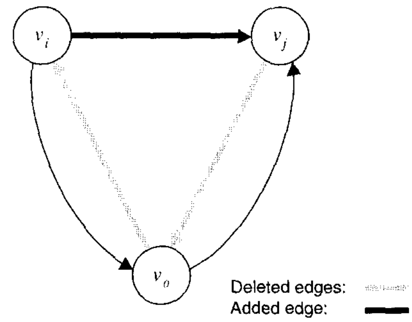

Fig. 2. Merging two routes in the savings algorithm.

The classical *savings algorithm* applies to problems for which the number of vehicles is not determined a priori. It starts with n vehicle routes, each containing the depot and a single customer. At each step, two routes are merged according to the largest achievable saving. The method can be outlined as follows.

Step 1. Compute the savings $s_{ij} = c_{i0} + c_{0j} - c_{ij}$, for $i, j = 1, 2, \ldots, n$. Create n vehicle routes (v_0, v_i, v_0) for $i = 1, 2, \ldots, n$.

Step 2. Order the savings in a non-increasing fashion.

Step 3. Consider two vehicle routes containing edges (v_0, v_i) and (v_0, v_j), respectively. If $s_{ij} > 0$, tentatively merge these routes by introducing edge (v_i, v_j) and by deleting edges (v_0, v_i) and (v_0, v_j). Implement the merge of the resulting route if feasible. Repeat this step until no improvement is possible (see Fig. 2).

The myopic aspect of this procedure, and the fact that it tends to produce a number of circumferential routes make it a relatively uninteresting alternative. However, several improvements suggested over the years tend to reduce these effects. One is to construct several routes in parallel (Altinkemer and Gavish, 1991), another is to weigh the term c_{ij} by a user controlled positive parameter α (Yellow, 1970), others are the use of sophisticated data structures to speed up the computations (Nelson et al., 1985).

The sweep algorithm works on planar problems for which the number of vehicles is not fixed. The method is easily implemented using an arbitrary ray first drawn from the depot. Customers v_i are represented by their polar coordinates (θ_i, ρ_i), where θ_i is the angle v_i makes with the depot and the ray, and ρ_i is the length of the segment $v_0 - v_i$. Assume all customers are ranked in increasing order of their polar angle.

Step 1. Select an unused vehicle k.

Step 2. Starting from the unrouted vertex with the smallest angle, assign the next customer (with the next largest angle) to vehicle k, as long as vehicle capacity is not exceeded. If unrouted vertices remain, go to step 1.

Step 3. Optimize each route separately by means of a heuristic for the Traveling Salesman problem (TSP) (Laporte, 1992a).

To illustrate how the algorithm works, consider the following ten customer example. The vehicle capacity is 10. Customers and their demands are given in Table 1. In Fig. 3, these customers are shown in a plane. Each customer v_i with demand q_i is represented by a pair (v_i, q_i). The initial ray is given by the full line; each

Table 1
Customers and demands

Customer	Demand	Customer	Demand
1	2	6	6
2	4	7	3
3	3	8	3
4	1	9	2
5	5	10	7

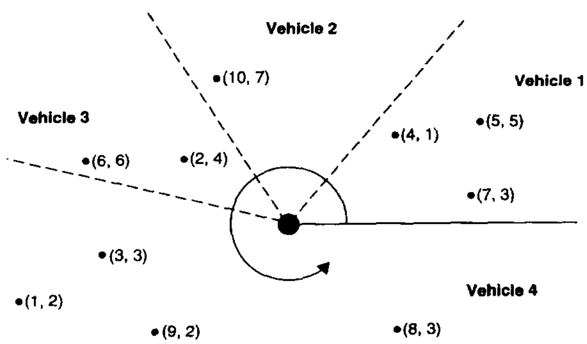

Fig. 3. Illustration of the sweep algorithm.

dashed line corresponds to a vehicle. In this example, the algorithm would create four vehicle routes with customer sets $\{4,5,7\}$, $\{10\}$, $\{2,6\}$, $\{1,3,8,9\}$.

In recent years, several *metaheuristics* have been proposed for the approximate resolution of a number of combinatorial optimization problems, including the VRP. Metaheuristics are general methods that guide subordinate heuristics in order to perform a clever search of the solution space. These include simulated annealing, tabu search and genetic algorithms (for recent surveys, see Pirlot, 1992; Reeves, 1993; Osman and Kelly, 1996). Tabu Search (TS) was introduced in the mid-eighties by Glover (1986) and Hansen (1986). It has been applied successfully to the VRP by a number of authors (see, e.g., Taillard, 1993; Gendreau et al., 1994; Rochat and Taillard, 1995). Typically, the method starts from an initial solution obtained by means of any heuristic. At a given iteration, it explores neighbor solutions, i.e., solutions that can be reached from the current solution by performing a given type of operation. It then moves to the best neighbor and the process is repeated until a termination criterion is met. Since successive solutions do not necessarily improve upon one another (as is the case in classical local search methods), solutions that were recently examined are inserted in a constantly updated *tabu list*, i.e., a list of forbidden solutions. Several implementation devices such as diversification and intensification are now commonly employed (see Glover and Laguna, 1993, for an overview). Here follows a short and simplified description of TABUROUTE, a tabu search algorithm designed by Gendreau et al. (1994) for the VRP with capacity and maximal route length restrictions.

Step 1. Construct a giant tour over all vertices. (Here the GENIUS TSP heuristic due to Gendreau et al. (1992) is used). Then break the tour into feasible VRP routes. Denote this solution by x and its cost by $F(x)$. Set the tabu list $T := \emptyset$.

Step 2. Attempt to improve the solution by applying the following search procedure to a subset W of $V \setminus \{v_0\}$ and by then selecting the best solution. Define $N(x)$, the neighborhood of x, as the set of all solutions that can be reached by inserting a vertex of W into a route containing one of its closest neighbors, using the GENI procedure (Gendreau et al., 1992). If $N(x) \setminus T = \emptyset$, go to step 3. Otherwise, identify a least cost solution y in $N(x) \setminus T$ and set $x := y$. Update the best known solution.

Step 3. If a preset number of iterations has been reached, go to step 4. Otherwise, update T and go to step 2.

Step 4. Attempt to improve the current solution by applying the search procedure to a larger set W, and allowing fewer iterations. If $N(x) \setminus T = \emptyset$, go to step 5. Otherwise, identify a least cost solution y in $N(x) \setminus T$ and set $x := y$. Update the best known solution.

Step 5. If the maximum number of iterations has been reached, stop. Otherwise, update T and go to step 4.

TABUROUTE has been applied to fourteen benchmark problems with $50 \leq n \leq 199$, taken from Christofides et al. (1979). It always produces better solutions than simpler heuristics, at the expense of increased computation time. The more recent TS implementation by Rochat and Taillard (1995) yields best known solutions on all of the fourteen test problem.

Whether one should use a rather involved heuristic such as TABUROUTE, as opposed to easy-to-program heuristics such as the savings method or the sweep algorithm, depends on the context. If quick and dirty solutions are required, the latter methods are probably appropriate. But if large sums of money are at stake, it may worth investing in more sophisticated heuristic in order to obtain a more economic solution.

4. Operational issues and models

The ultimate goal of any transportation firm is to make profits and improve, or at least maintain its competitive position. To this end, strategic and tactic plans can be drawn up to guide operations, but the operational capabilities of the firm will ultimately determine its performance. Hence, models and tools aimed at assisting decision making at an operational level are an important component of a comprehensive decision support system.

There are many different issues which have to be addressed at the operational level in order to ensure that demand is satisfied within the required service criteria and the resources of the carrier are efficiently used. Most of these issues have to consider the *time* factor: a response may be required in real or near-real time (e.g., assigning an empty container to a customer request), today's decisions may have a significant impact on future decisions and performances (e.g., if empty rail cars are not repositioned, in a few days equipment will be idle in some yards while it will not be possible to satisfy demands at some other terminals), schedules may be needed for vehicles and crews, time conditions are imposed on operations (e.g., a container has to arrive in time to be loaded on the departing ship or a truck has to pick up a load within a specified time window), etc. In other type of operations the very notion of a planned solution does not make sense. Consider, for example, courier services operating in urban contexts. Here, drivers often start their working day with a short list of collection or delivery requests. As the day progresses, new requests will be communicated to the drivers and will have to be incorporated into their routes.

Most models traditionally used in transportation planning use as their input known static data. For examples, VRP models operate with given customer demands and travel times, tactical planning formulations consider aggregated forecast demand data as "known", etc. However, the real world in which these models are implemented is in a constant state of change and solutions cannot always be implemented as planned. If traffic is slower than predicted, vehicles may arrive late at the customers' locations and at the depot. If supplies are larger than forecast, vehicles may become full prematurely and it may become necessary to pay penalties (e.g., pay drivers at overtime rate for part of the day) or to implement recourse actions (e.g., when a delivery vehicle becomes empty, it could go back to the depot to replenish and resume deliveries starting at the next point in its planned route).

Consequently the *dynamic* aspect of operations is further compounded by the *stochasticity* inherent in the system, that is by the set of uncertainties that are characteristic of real-life management and operations. Increasingly, these characteristics are reflected into the models and methods aimed at operational planning and management issues, as illustrated in the following.

4.1. Dynamic modeling to support carrier operations

The problems most often encountered in the operational planning and management of transportation carriers include the following (see also Powell et al., 1995b, and references therein):

(1) *Scheduling of services.* A tactical load plan will often indicate which service to offer and, eventually, how often to run it over the planning period. Often, however, service will be offered according to a *schedule* that indicates the time of departure at the origin and the time of arrival at the destination, as well as the time and length of stops at intermediate terminals, where appropriate. The schedule may be fixed (e.g., airlines) or it may indicate a departure interval (e.g., a truck of a LTL motor carrier leaves the terminal between 7 and 9 p.m.) in which case customers and other operating units of the company will be given a cut-off time for their loads to make the service in time. Passenger transportation usually takes place according to fixed schedules and so usually is air cargo. LTL trucking is generally operated according to interval schedules. Traditionally, North American railways did not operate scheduled services or the schedule represented only a goal, or a general idea of how the trains were run. Currently, however, there is a trend towards totally or partially scheduled services to increase service levels and the competitiveness of the firm. In Europe, there is a longer tradition of operating scheduled rail services (with varying degrees of success), even of including booking schemes for freight trains (see, for example, Joborn, 1995).

(2) *Empty vehicle distribution or repositioning.* The imbalance between freight demand and supply is reflected in the fact that at any point in time there are terminals with a surplus number of vehicles of a certain type, while some other terminals show a shortage. Then, vehicles have to be moved empty (or additional loads have to be found) in order to bring them were they will be needed to satisfy known and forecast demand. Moving vehicles empty does not directly contribute to the profit of the firm but it is essential to its continuing operations. Consequently, one attempts to minimize empty movements within the limits imposed by the demand and service requirements. Empty vehicle distribution is a central component of planning and operations of many transportation firms, especially in the rail, container and LTL motor carrier industries. There is, therefore, a very rich literature addressing these issues. The Dejax and Crainic (1987) survey reviews contributions going back to the 60's and spans the whole spectrum of modeling approaches from simple static transport models to formulations that integrate the dynamic and stochastic characteristics of the problem (e.g., the work of Jordan and Turnquist, 1983). Significant research efforts continue to be directed toward this class of problems as illustrated by the recent contributions of Chih (1986), Adamidou et al. (1993), Joborn (1995), Turnquist (1994), and others.

(3) *Crew scheduling.* Crews have to be assigned to vehicles and convoys in order to support the planned operations. There are also numerous other issues related to manpower management such as the scheduling of reserve crews, terminal employees, maintenance crews, etc. A significant body of methodological and technological knowledge has been developed to deal with these issues, especially in the context of transit (bus and passenger rail) and airline transportation (see, for example, Barnhart and Talluri, 1997; Desrosiers et al., 1995). These methodologies were developed for applications where detailed schedules are known and adhered to. Consequently, although some efforts have been aimed in this direction (see, for example, Crainic and Roy, 1990, 1992), currently it appears that better results can be achieved by applying the class of methodologies used to dynamically allocate resources to tasks.

(4) *Allocation of resources.* Many operational problems may be viewed as dynamically allocating resources to tasks (Powell, 1995). For example, one has to allocate empty vehicles (trailers, railcars, etc.) to the appropriate terminals, motive power (tractors, locomotives, etc.) to services, crews to movements or services, customer loads to driver-truck combinations, empty containers from depots to customers and returning containers from customers to depots, etc. This is an extremely rich field both for research and development and for applications. Dynamic and stochastic network formulations have been and continue to be extensively studied. This has resulted in important modeling and algorithmic results; see, for example, Powell (1988), Frantzeskakis (1990) and Frantzeskakis and Powell (1990), Powell and Cheung (1994a, b, 1996), Crainic et al. (1993a), etc. A number of these results have been transferred to industry (Powell et al., 1992, for example). The interested reader should consult the excellent synthesis and review by Powell et al. (1995b) and the numerous references quoted in that work. The recently proposed Logistics Queueing Net-

works methodology (Powell et al., 1994, 1995a; Carvalho, 1996) appears, however, to offer an even more interesting framework for a wide variety of real situations which may be efficiently represented and solved.

It is not possible in the present paper to provide an exhaustive review of the models and methods developed for the issues just described. We encourage the interested reader to consult the documents mentioned. In the following, we briefly illustrate two such modeling approaches.

The first example concerns the issue of determining the dynamic service network of a carrier which operates according to more or less fixed schedules. Rail companies and less-than-truckload motor carriers are examples of such firms. The main issues are similar to those of the tactical planning problem described in Section 3.1, with one major difference. While in tactical planning one is concerned with the *where* and *how* issues (selecting services of given types and traffic routes between spatial locations), here one is interested above all in *when* issues: when to start a given service, when the vehicle arrives at destination or at an intermediary terminal, when is the traffic delivered, etc.

A very simple instance of such a problem is illustrated in Fig. 4. For the sake of simplicity, assume that for the service network illustrated in Fig. 1, one has already decided that out of terminal A only services S1, S2 and S3 will be offered. One must then determine the departure schedule for these services over the next periods, as well as the corresponding traffic routing. Part of the dynamic time-space network that may be used to support these decisions is shown in Fig. 4. Here, each terminal is drawn for each of the periods considered (six, in this example). Each service is then drawn at each of its possible departure period, and its route in space and time is shown. Hence, service S1 departing in period 1 arrives at terminal D one period later, while service S2 takes two periods to reach the same destination. A more complex route is shown for service S2 which, departing at the same period as the other two, arrives at terminal C one period later, waits there for one period for loading, unloading and consolidation operations, to finally arrive at terminal D in period 4. The same *dynamic service arcs* are also drawn starting at the other periods. Other dynamic arcs are the *holding* links which represent waiting for traffic in terminals; the arcs from node A1 to node A2 is such

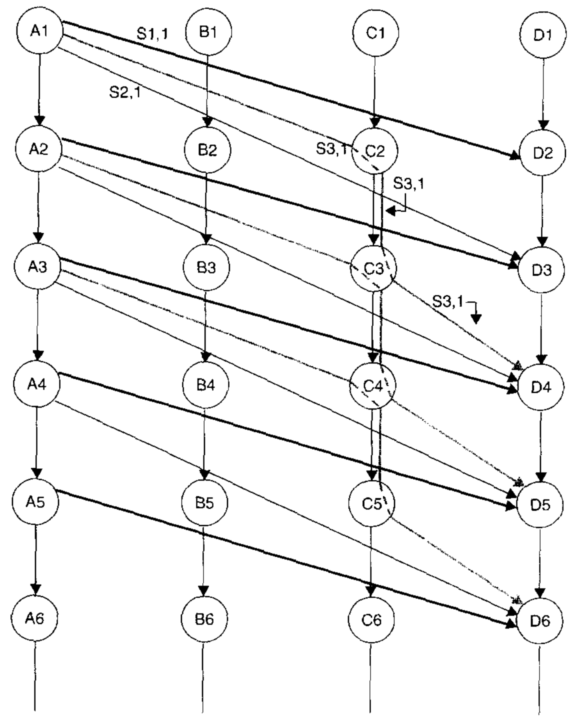

Fig. 4. Example of dynamic service design network.

a holding arc. Different holding arcs could be used to represent various terminal activities. Note that, in order not to overload the picture, we have not shown all the dynamic arcs, nor all terminals, nor all the other arcs required to model the entry and exit of demand.

Models for this class of problems may be written similarly to the network design formulations of Section 2 (Eqs. (11)–(16)). Here, one associates to each dynamic service arc an integer variable equal to 1 if the corresponding departure is chosen, and to zero otherwise. Other variables capture the flow of traffic on the services and through terminals. Additional constraints, such as terminal and service capacities or the time windows imposed on the delivery of goods at destination, further complicate the problem (for a complete formulation, see Farvolden and Powell, 1991, 1994).

As already indicated, the problem of allocating limited resources to requests and tasks is one of the most typical and critical issues relative to operating a transportation system (as well as most industrial and service systems, for that matter). Repositioning equipment, engines, tractors, rail cars, trailers ships, etc., to respond to sure and forecast demands, allocating crews to transportation services or trucks and containers to loads, all these types of problems have several common characteristics:

(1) some future demands are known, but most can only be forecast and unpredictable requests may happen;

(2) many requests materialize in real or quasi real time and have to acted upon in relatively short time;

(3) once a resource is allocated to an activity it is no longer available for a certain duration (whose length may be subject to variations as well);

(4) once a resource becomes available again, it is often in a different location than its initial one;

(5) the value of an additional unit of a given resource at a location greatly depends on the total quantity of resources available (which are determined by previous decisions at potentially all terminals in previous periods) and the current demand, etc.

One may represent such issues by an activity graph similar to the one displayed in Fig. 5. Here, the operations of a simple four terminal system are schematically drawn for a certain length of time (one week, for example), arbitrarily divided into three periods (half days, for example). At each terminal, there are a number of vehicles (containers, for example) which are currently available to satisfy customers requests during the current period and in future ones. Customer demands have precise characteristics, such as origin and destination of movement, pick up and delivery dates (with time windows, eventually), etc. At any period,

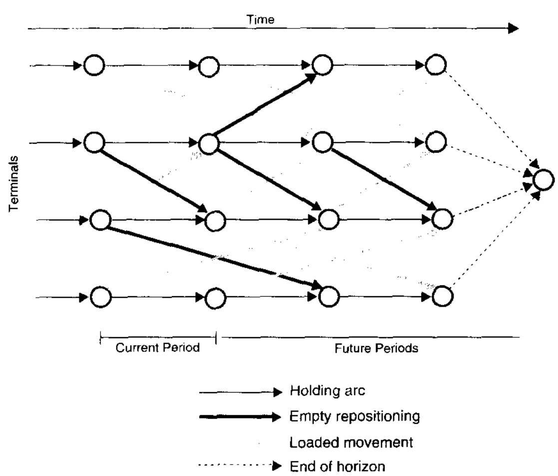

Fig. 5. Space-time diagram for dynamic resource allocation.

a vehicle may be assigned to a customer demand at the current period and location, or it may be moved to another location to satisfy a known future request, or it may be held at the current location or moved empty to another location in preparation for future, forecast demands. The network of Fig. 5 illustrates these possibilities and decisions.

Accepting requests and performing the corresponding movements implies expenses and generates revenues. Some models address the issue of whether the request is profitable with respect to the operation of the system and should therefore be accepted (Powell et al., 1992, for example). Repositioning empty vehicles does not generate any immediate revenues. One may be ready to incur these expenses, however, in the hope that, as a consequence, vehicles will be adequately posted to take advantage of future (known, forecast or only guessed at) opportunities. Refused requests represent lost business opportunities, while accepted but unsatisfied ones generally result in penalties.

A classical modelling approach for this class of problems is to consider the entire planning horizon with the objective of maximizing the *total system profit* computed as the sum of the profit resulting from decisions taken for the current period, plus the expected profit over future periods. The usual constraints (e.g., satisfy demands, do not use more than the number of available vehicles, etc.) apply. When the state of the system and its environment in future periods is known, or assumed to be known, the resulting formulation is deterministic and is often written as a network optimization model with additional constraints.

The major difficulty with this approach becomes apparent when the uncertainties in future demands (as well as, eventually, uncertainties related to performing the operations) are explicitly considered. In this case, decisions taken "now" for future periods cannot be based on "sure" data, but only on estimations of how the system will evolve, which demand will materialize, etc. From a mathematical programming point of view, random variables are used to represent the stochastic elements and decisions in future periods. Consequently, the expectation of future profits that appears in the objective function of the model becomes a very complex recursive stochastic equation where the statistical expectation of the total profit has to be computed over all possible realizations of all random variables.

To address this complex issue, the model generally takes the form of a recourse formulation. Such formulations are based on the idea that today's decisions are taken within today's deterministic context but using an "estimation" of the variability of the random factors, and that their consequences are reflected in later decisions. The recourse represents these later decisions which have to be taken to adjust the initial policies once the actual realization of the random variables is observed. In the simplest possible recourse formulation, called *simple recourse*, it is assumed that one does not attempt to change the decisions but pays a penalty when the observed value of a random variable is different from the estimation. More complex formulations, such as nodal and network recourse (see Powell and Frantzeskakis, 1994, Powell and Cheung, 1994a, and references therein), attempt to evaluate the possible modifications to the initial decisions, and the impact on the total expected profit. An excellent analysis of the application of these approaches to the dynamic fleet management problems for truckload motor carriers, as well as a discussion of the merits and difficulties of stochastic formulations, may be found in Powell (1996).

These formulations, which are generally difficult to solve, also make use of various criteria to discretize, aggregate and end time. For example, in Fig. 5, the theoretically infinite future planning and operations horizon has been reduced to three periods. When the recourse formulation is solved, the periods could be further aggregated, all future ones being considered as one; this corresponds to a two-period formulation, as opposed to n-period, otherwise. Then, in actual application, the models are used in a rolling horizon environment where, as time advances, a new period is added at the end of the horizon. An important issue is then how to approximate what happens in all the periods beyond the artificially fixed end of the horizon, and how to integrate this approximation into the recourse function. Powell et al. (1995b) present an excellent review of this class of formulations.

A different approach recently championed by Powell (1995; see also Powell et al., 1995a; Carvalho, 1996) addresses resource allocation problems as *Logistic Queueing Networks*. In this case, at each node of the time-space diagram there are two queues: one of resources and one of tasks requesting resources. Fig. 6 illustrates a possible configuration for the first

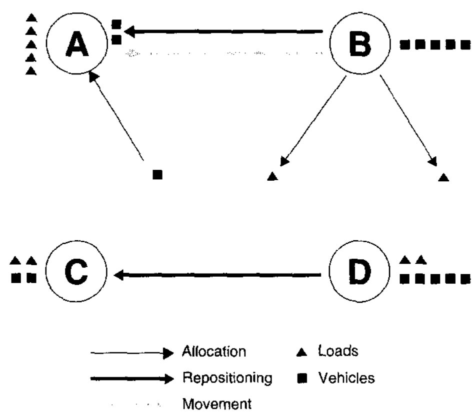

Fig. 6. Example of resource allocation.

period of a four terminal case. The basic idea is that in order to evaluate the worth of allocating a vehicle to a loaded movement, one has to know, or evaluate, not only the operating costs and the price of the load, but also the value of the vehicle to the destination terminal. Then, at each terminal and for each possible destination, these values have to be computed in order to decide on the allocations. Similar values have to be computed for empty movements. Various other considerations, such as time windows, labor restrictions, substitutions, etc., may also be included in the model. The general solution approach makes a series of forward (to allocate vehicles) and backward (to evaluate vehicle values at nodes) passes, until "convergence" is ensured. Several issues have to be addressed before implementing such an approach, but the initial results appear very promising.

4.2. Capacitated routing with uncertainties

We now describe the two approaches commonly used for dealing with stochasticity in transportation planning, in general, and vehicle routing, in particular: a priori (Bertsimas et al., 1990) optimization and real-time optimization.

In a priori optimization, problems are formulated within the framework of stochastic programming and are modeled in two stages. In a first stage, a planned or a priori solution is designed. The values taken by some random variables are then disclosed and, in a second stage, a recourse action is taken, such as paying drivers overtime or returning to the depot to replenish. A stochastic program is usually modeled as a chance constrained program (CCP) or as a stochastic program with recourse (SPR). In CCP, the planned solution is designed such that its probability of failure lies below a certain acceptable threshold. In SPR, the first stage solution must be such that the expected cost of the second stage solution is minimized. SPR is more realistic than CCP in that it works on the true expected cost of a solution, but it is also more difficult to develop solution methodologies for this case.

Before proceeding further, we illustrate these two solution concepts in the context of the CVRP. Consider the simple three customer problem depicted in Fig. 7. Customers 1 and 2 have a deterministic demand of 4, and the demand of customer 4 is 2 with probability 0.9 and 4 with probability 0.1. Let the vehicle capacity be equal to 10. For the CCP model, assume the probability of route failure is $\alpha = 0.05$. Then, an optimal solution is to use two vehicles and thus ensure that route failure will never occur. To implement an SPR solution, one must first define an appropriate recourse policy. Here the policy is for the delivery vehicle to follow its planned route until its capacity becomes empty. In such a case, it returns to the depot to replenish and resumes its deliveries along the planned route. The optimal planned route is $(0,1,2,3,0)$ and has a cost of $2+2\sqrt{2}$. If the demand of customer 3 is 2, this is also the cost of the second stage solution. If the demand of customer 3 is 4, then upon arriving at that customer, the vehicle must return to the depot to replenish, and go back to customer 3 to carry out its last delivery. Thus, the expected cost of that solution is $2+2\sqrt{2}+0.1(2\sqrt{2}) = 2+2.2\sqrt{2}$. Note that in this case, the reverse first stage solution $(0,3,2,1,0)$ yields the same expected cost but, as shown by Dror and Trudeau (1986), this is not always the case. Figs. 8 and 9 illustrate the solutions of the two models.

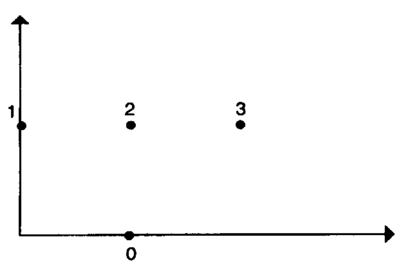

Fig. 7. Stochastic VRP-customer locations.

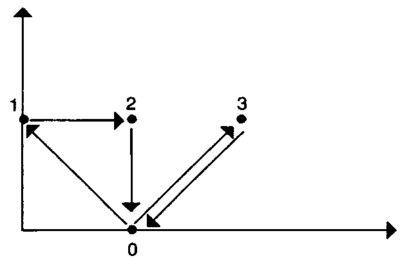

Fig. 8. Optimal CCP solution using two vehicles.

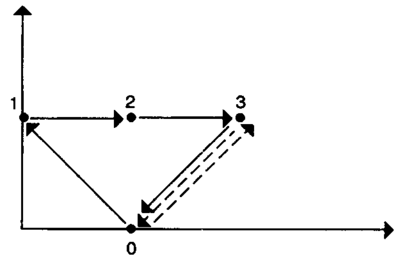

Fig. 9. Optimal SPR solution using one vehicle and a recourse strategy.

How does one solve stochastic VRPs in practice? Due to the extreme complexity of such problems, most authors resort to heuristics, typically adaptations of known heuristics for the deterministic VRP. Thus Dror and Trudeau (1986) use a modification of the Clarke and Wright (1964) savings algorithm. More recently, Gendreau et al. (1996a) have devised a tabu search heuristic for a more general version of the stochastic VRP in which customers are present with some probability (and are skipped if they are not present) and have stochastic demands. The method uses a proxy objective function to avoid computing at each candidate solution the true expected solution cost. Until quite recently, it was difficult to assess the quality of solutions found by heuristics since there were no good lower bounds and no exact solution available. Some progress is, however, being made in these areas. A new type of exact algorithm, called the Integer L-shaped method, was recently proposed by Laporte and Louveaux (1993) for a wide class of stochastic integer programs with recourse. Briefly the method solves in a branch-and-cut fashion the first stage program containing a lower bound on the cost of recourse. At any feasible solution, the true cost of recourse is computed. If the solution is dominated, fathoming occurs; otherwise, the branching process is reinitiated from that solution until dominance occurs, or until the cost of the relaxed problem coincides with the cost of the true solution. This method was successfully applied to a family of SVRPs by Gendreau et al. (1995). For some values of the parameters it was able to solve to optimality VRPs with stochastic demands involving up to 70 customers. Smaller sizes were attained when both the customer presence and their demands were stochastic. Using these results, the authors were able to demonstrate that on instances for which an optimal solution was available, their tabu search algorithm yielded solutions that were optimal more than 89% of the time, and the average deviation from optimality was only 0.38%. For a survey on stochastic VRP, interested readers are referred to Gendreau et al. (1996b).

The type of stochastic VRP just described is relatively basic in that information on demands arrives only when drivers reach the customers' locations and then a very simple type of recourse action is applied: return to the depot or carry on along the planned route. More sophisticated types of recourse are available such as preventive return trips to the depot could be made even if route failure has not occurred, in order to avoid a potentially longer penalty later on. Such preventive breaks could take place, for example, when the vehicle is near the depot. A more sophisticated type of recourse action would be to obtain information on the demand of the following customer whenever the vehicle arrives at a location and decide at that stage whether a preventive break is necessary. At the other end of the spectrum, one could envisage a situation where information on customers' demands is continuously transmitted to the driver as he follows his route, and where the remaining portion of the route is reoptimized in real time. This requires, of course, the availability of efficient and reliable optimization algorithms.

Some of these scenarios are now a step closer to reality with the dissemination of new information technologies (see, e.g., Psaraftis, 1995; Powell et al., 1995b), which brings us to the problem of real-time vehicle dispatching.

As mentioned earlier, it is now common for courier firms to provide instructions in real time to their drivers, upon reception of new requests. What distinguishes these problems from the SVRP is that the set of potential customers is to all practical purposes infinite. In such contexts one can only hope to devise good operating policies. Here, simulation can be used to generate requests and to compare alternative policies.

In a recent paper, Shen et al. (1995) described a dispatching system for a courier service company. Throughout the day, customers who have a letter to deliver phone the dispatching office and state their preferences (a time window) with regards to pick up and delivery, e.g. the letter should be picked up within the next three hours. The dispatcher must assign that request to one of the available vehicles, taking into account the current location and current planned route of each driver, locations of pick up and delivery points of the new request, the distances and travel points between points, etc. The authors have developed a learning system, based on neural networks, to assign new requests to drivers. Decisions taken by the system are then compared to those made by a professional dispatchers, which enables the system to readjust its decision rules. The authors found a good fit between the decisions made by the system and those made by the dispatchers. They conclude that neural network methodology can be successfully applied to such a system provided sufficient information can be obtained in real time on drivers' locations.

The development of real-time routing systems is still in its early stages but is quickly emerging as a rich research area. There is no doubt that its growth is closely link to that of new information technologies. However, technology alone is not sufficient to ensure success. The need remains to develop quick and efficient algorithms capable of providing good solutions in real time.

5. Conclusions and perspectives

Freight transportation lies at the heart of our economic life. In industrialized countries, it accounts for a significant share of the gross national product. In developing countries, it is the essential ingredient of sustainable development. With free trade zones emerging in several parts of the world and with the globalization of the economic system, transportation will in all likelihood play an even more major role in years to come.

The trend towards larger, more integrated and more efficient transportation systems is likely to remain and should create the need for better planning at the strategic, tactical and operational levels. Transportation planning is undoubtedly one of the great success stories of Operations Research. Classical O.R. models and algorithms have consistently proved highly suitable for the solution of complex transportation problems at all planning levels and involving just about any mode. Several planning tools developed by operations researchers are now widely available and are routinely used by transportation planners.

The rapid growth in the development of computerized planning systems, witnessed over the last fifteen to twenty years, is mostly due to three factors: the design of more realistic models, the development of more powerful algorithms and the availability of more performing computers and of friendlier user interfaces. These advances have had a direct impact on the realism, complexity and size of planning models in the field of freight transportation, and on their level of acceptance by planners and operators.

The growth of transportation planning methods will likely continue to be driven by the same factors in years to come. For example, major research efforts are being devoted to the design of models for dynamic and stochastic problems, bi-level programming enables the representation of competition between players, etc. Major developments are also taking place in the artificial intelligence-related area of metaheuristics such as tabu search, genetic algorithms, neural networks, etc. These have already given a new impetus to the whole area of global optimization and have lead to a rethinking of the entire field of heuristics. These developments, coupled with the growth of parallel methods, mean that in the near future larger and more complex problems should be amenable to analysis and optimization. In particular, significant advances should be expected in the areas of dynamic, stochastic and real-time programming, central to so many transportation systems.

Acknowledgements

Funding for this project has been provided by the Natural Sciences and Engineering Council of Canada and by the Fonds F.C.A.R. of the Province of Quebec. We would like to thank Lucie Larin for her assistance with the editing of the manuscript and the professional drawing of the figures. This article has been adapted from a chapter in *Civil Engineering Systems: An Advanced Applications Text*, which will be published by John Wiley & Sons in 1997.

References

Adamidou, E.A., Kornhouser, A.L., and Koskosidis, Y.A. (1993), "A game theoretic/network equilibrium solution approach for the railroad car management problem", *Transportation Research B* 27B, 237–252.

Ahuja, R.K., Magnanti, T.L., and Orlin, J.B. (1993), *Network Flows - Theory, Algorithms, and Applications*, Prentice-Hall, Englewood Cliffs, NJ.

Altinkemer, K., and Gavish, B. (1991), "Parallel savings based heuristic for the delivery problem", *Operations Research* 39, 456–469.

Ash, L.L. (1993), "Leadership and the North American trucking industry", in: B.G. Bison, ed., *A Look Back from the Year 2000*, 28th Annual Meeting of the Canadian Transportation Research Forum, 30–40.

Assad, A.A. (1980), "Models for rail transportation", *Transportation Research A* 14, 205–220.

Assad, A.A. (1988), "Modeling and implementation issues in vehicle routing", in: B.L. Golden and A.A. Assad, eds., *Vehicle Routing: Methods and Studies*, North-Holland, Amsterdam, 7–45.

Balakrishnan, A., Magnanti, T.L., and Wong, R.T. (1989), "A Dual-ascent procedure for large-scale uncapacitated network design", *Operations Research* 37/5, 716–740.

Barnhart, C., and Kim, D. (1995), "Routing models and solution procedures for regional less-than-truckload operations", *Annals of Operations Research* 61, 67–90.

Barnhart, C., and Talluri, K. (1997), "Airlines operations research", in: A.E. McGarity and C. ReVelle, eds., *Civil and Environmental Engineering Systems: An Advanced Applications Text*, Wiley, New York, NY.

Bertsimas, D.J., Jaillet, P., and Odoni, A.R. (1990), "A priori optimization", *Operations Research* 38/1, 1019–1033.

Bodin, L.D., Golden, B.L., Assad, A.A., and Ball, M.O. (1983), "Routing and scheduling of vehicles and crews: The state of the art", *Computers and Operations Research* 10/2, 63–211.

Braklow, J.W., Graham, W.W., Hassler, S.M., Peck, K.E., and Powell, W.B. (1992), "Interactive optimization improves service and performance for yellow freight system", *Interfaces* 22/1, 147–172.

Button, K.J. (1993), *Transport Economics*, Edward Elgar.

Chen, B., and Harker, P.T. (1990), "Two moments estimation of the delay on a single-track rail line with scheduled traffic", *Transportation Science* 24/4, 261–275.

Chih, K.C.-K. (1986), "A real time dynamic optimal freight car management simulation model of the multiple railroad, multicommodity temporal spatial network flow problem", Ph.D. Thesis, Princeton University.

Christofides, N., Mingozzi A., and Toth, P. (1979), "The vehicle routing problem", in: N. Christofides, A. Mingozzi, P. Toth, and C. Sandi, eds., *Combinatorial Optimization*, Wiley, New York, NY, 315-338.

Clarke, G., and Wright, J.W. (1964), "Scheduling of vehicles from a central depot to a number of delivery points", *Operations Research* 12, 568-581.

Cordeau, J.-F., and Laporte, G. (1996), "Le transport des marchandises en quelques chiffres", *Routes et Transports* 25, 43-44.

Cornuéjols, G., and Harche, F. (1993), "Polyhedral study of the capacitated vehicle routing problem", *Mathematical Programming* 60, 21-52.

Crainic, T.G. (1982), "Un modèle de plannification tactique pour le transport ferroviaire des marchandises", Ph.D. Thesis, Université de Montréal.

Crainic, T.G. (1984), "A comparison of two methods for tactical planning in rail freight transportation", in: J.P. Brans, ed., *Operational Research'84*, North-Holland, Amsterdam, 707-720.

Crainic, T.G. (1988), "Rail tactical planning: Issues, models and tools", in: L. Bianco and A. La Bella, eds., *Freight Transport Planning and Logistics*, Springer-Verlag, Berlin, 463-509.

Crainic, T.G., and Nicolle, M.-C., (1986), "Planification tactique du transport ferroviaire des marchandises: quelques aspects de modélisation", in: *Actes du Premier Congrès International en France de Génie Industriel*, CEFI-AFCET-GGI, Paris, Tome 1, 161-174.

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

Crainic, T.G., and Roy, J. (1988), "O.R. tools for tactical freight transportation planning", *European Journal of Operational Research* 33/3, 290-297.

Crainic, T.G., and Roy, J. (1990), "Une approche de recouvrement d'ensembles pour l'établissement d'horaires des chauffeurs dans le transport routier de charges partielles", *R.A.I.R.O. - Recherche opérationnelle* 24/2, 123-158.

Crainic, T.G., and Roy, J. (1992), "Design of regular intercity driver routes for the LTL motor carrier industry", *Transportation Science* 26/4, 280-295.

Crainic, T.G., J.-A. Ferland, and Rousseau, J.-M. (1984), "A tactical planning model for rail freight transportation", *Transportation Science* 18/2, 165-184.

Crainic, T.G., Dejax, P.J., and Delorme, L. (1989), "Models for multimode multicommodity location problems with interdepot balancing requirements", *Annals of Operations Research* 18, 279-302.

Crainic, T.G., Florian, M., Guélat J., and Spiess, H. (1990a), "Strategic planning of freight transportation: STAN, an interactive-graphic system", *Transportation Research Record* 1283, 97-124.

Crainic, T.G., Florian, M., and Léal, J.-E. (1990b), "A model for the strategic planning of national freight transportation by rail", *Transportation Science* 24/1, 1-24.

Crainic, T.G., Gendreau, M., and Dejax, P.J. (1993a), "Dynamic stochastic models for the allocation of empty containers", *Operations Research* 43, 102-302.

Crainic, T.G., Gendreau, M., Soriano, P., and Toulouse, M. (1993b), "A tabu search procedure for multicommodity location/allocation with balancing requirements", *Annals of Operations Research* 41, 359-383.

Crainic, T.G., Gendreau, M., and Farvolden, J.M. (1996), "Simplex-based tabu search for the multicommodity capacitated fixed charge network design problem", Publication CRT-96-07, Centre de recherche sur les transports, Université de Montréal.

Daganzo, C.F. (1991), *Logistics Systems Analysis*, Lecture Notes in Economics and Mathematical Systems 361, Springer-Verlag, Berlin.

Daskin, M.S. (1995), *Networks and Discrete Location*, Wiley, New York, NY.

de Carvalho, T.A. (1996), "A new approach to solving dynamic resource allocation problems", Ph.D. Thesis, Department of Civil Engineering and Operations Research, Princeton University.

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

Delorme, L., and Crainic, T.G. (1993), "Dual-ascent procedures for multicommodity location-allocation problems with balancing requirements", *Transportation Science* 27/2, 90-101.

Delorme, L., Roy, J., and Rousseau, J.-M,.(1988), "Motor-carrier operation planning models: A state of the art", in: L. Bianco and A. La Bella, eds., *Freight Transport Planning and Logistics*, Springer-Verlag, Berlin, 510-545.

Desrosiers, J., Dumas, Y., Solomon, M.M., and Soumis, F. (1995), "Time constrained routing and scheduling", in: M. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, eds., *Handbooks in Operations Research and Management Science 8. Network Routing*, North-Holland, Amsterdam, 35-139.

Dror, M., and Trudeau, P. (1986), "Stochastic vehicle routing problem", *Journal of Operational Research* 23, 282-235.

Eilon, S., Watson-Gandy, C.D.T., and Christofides, N. (1971), *Distribution management: Mathematical modelling and practical analysis*, Griffin, London.

Erlenkotter, D. (1978), "A dual-based procedure for uncapacitated facility location", *Operations Research* 26, 992-1009.

Farvolden, J.M., and Powell, W.B. (1991), "A dynamic network model for less-than-truckload motor carrier operations", Working Paper 90-05, Department of Industrial Engineering, University of Toronto.

Farvolden, J.M., and Powell, W.B. (1994), "Subgradient methods for the service network design problem", *Transportation Science* 28/3, 256-272.

Fisher, M.L. (1995), "Vehicle routing", in: M. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, eds., *Handbooks in Operations Research and Management Science 8. Network Routing*, North-Holland, Amsterdam, 1-33.

Florian, M. (1984), "An introduction to network models used in transportation planning", in: M. Florian, ed., *Transportation Planning Models*, North-Holland, Amsterdam, 137-152.

Florian, M. (1986), "Nonlinear cost models network models in transportation analysis", *Mathematical Programming Study* 26, 167-196.

Florian, M., and Hearn, D. (1995), "Networks equilibrium models and algorithms", in: M. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, eds., *Handbooks in Operations Research and Management Science 8. Network Routing*, North-Holland, Amsterdam, 485–550.

Frantzeskakis, L. (1990), "Dynamic networks with random arc capacities, with application to the stochastic dynamic vehicle allocation problem", Ph.D. Thesis, Department of Civil Engineering, Princeton University.

Frantzeskakis, L., and Powell, W.B. (1990), "A successive linear approximation procedure for stochastic dynamic vehicle allocation problems", *Transportation Science* 24/1, 40–57.

Friesz, T.L., and Harker, P.T. (1986), "Freight network equilibrium: A review of the state of the art", in: A.F. Daughety, ed., *Analytical Studies in Transport Economics*, Ch. 7, Cambridge.

Friesz, T.L., Gottfried, J.A., and Morlok, E.K. (1986), "A sequential shipper-carrier network model for predicting freight flows", *Transportation Science* 20, 80–91.

Gendreau, M., Hertz, A., and Laporte, G. (1992), "New insertion and postoptimization procedures for the traveling salesman problem", *Operations Research* 40/6, 1086–1094.

Gendreau, M., Hertz, A., and Laporte, G. (1994), "A tabu search heuristic for the vehicle routing problem", *Management Science* 40, 1276–1290.

Gendreau, M., Laporte, G., and Séguin, R. (1995), "An exact algorithm for the vehicle routing problem with stochastic demands and customers", *Transportation Science* 29, 143–155.

Gendreau, M., Laporte, G., and Séguin, R. (1996a), "A tabu search algorithm for the vehicle routing problem with stochastic demands and customers", *Operations Research* 44/3, 469–477.

Gendreau, M., Laporte, G., and Séguin, R. (1996b), "Stochastic vehicle routing", *European Journal of Operational Research* 88, 3–12.

Gendron, B., and Crainic, T.G. (1994), "Relaxations for multicommodity network design problems", Publication CRT-965, Centre de recherche sur les transports, Université de Montréal.

Gendron, B., and Crainic, T.G. (1995), "A branch-and-bound algorithm for depot location and container fleet management", *Location Science* 3/1, 39–53.

Gendron, B., and Crainic, T.G. (1996), "Bounding procedures for multicommodity capacitated network design problems", Publication CRT-96-06, Centre de recherche sur les transports, Université de Montréal.

Gillett, B., and Miller, L. (1974), "A heuristic algorithm for the vehicle dispatch problem", *Operations Research* 22, 340–349.

Glover, F. (1986), "Future paths for integer programming and links to artificial intelligence", *Computers & Operations Research* 1/3, 533–549.

Glover, F. (1989), "Tabu search - Part I", *ORSA Journal on Computing* 1/3, 190–206.

Glover, F. (1990), "Tabu search - Part II", *ORSA Journal on Computing* 2/1, 4–32.

Glover, F., and Laguna, M. (1993), "Tabu search", in: C.R. Reeves, ed., *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell Scientific Publications, Oxford, 70–150.

Golden, B.L., and Assad, A.A., eds. (1988), *Vehicle Routing: Methods and Studies*, North-Holland, Amsterdam.

Golden, B.L., and Wasil, E.A. (1987), "Computerized vehicle routing in the soft drink industry", *Operations Research* 35, 6–17.

Guélat, J., Florian, M., and Crainic, T.G. (1990), "A multimode multiproduct network assignment model for strategic planning of freight flows", *Transportation Science* 24/1, 25–39.

Haghani, A.E. (1989), "Formulation and solution of combined train routing and makeup, and empty car distribution model", *Transportation Research B* 23/6, 433–431.

Hansen, P. (1986), "The steepest ascent mildest descent heuristic for combinatorial optimization", in: *Congress on Numerical Methods in Combinatorial Optimization*, Capri, Italy.

Harker, P.T., and Friesz, T.L. (1986a), "Prediction of intercity freight flows I: Theory", *Transportation Research B* 20/2, 139–153.

Harker, P.T., and Friesz, T.L. (1986b), "Prediction of intercity freight flows II: Mathematical formulations", *Transportation Research B* 20/2, 155–174.

Harker, P.T., and Hong, S. (1990), "Two moments estimation of the delay on a partially double-track rail line with scheduled traffic", *Journal of the Transportation Research Forum* 31/1, 38–49.

Joborn, M. (1995), "Empty freight car distribution at Swedish Railways - Analysis and optimization modeling", Ph.D. Thesis, Division of Optimization, Department of Mathematics, University of Linköping.

Jordan, W.C., and Turnquist, M.A. (1983), "A stochastic dynamic network model for railroad car distribution", *Transportation Science* 17, 123–145.

Jovanović, D. and Harker, P.T. (1990), "Tactical scheduling of rail operations: The SCAN I decision support system", *Transportation Science* 25/1, 46–64.

Keaton, M.H. (1989), "Designing optimal railroad operating plans: Lagrangian relaxation and heuristic approaches", *Transportation Research B* 23, 415–431.

Keaton, M.H. (1991), "Service-cost tradeoffs for carload freight traffic in the U.S. rail industry", *Transportation Research A* 25/6, 363–374.

Keaton, M.H. (1992), "Designing optimal railroad operating plans: A dual adjustment method for implementing Lagrangian relaxation", *Transportation Science* 26/3, 263–279.

Kruskal, J.B. (1956), "On the shortest spanning tree of graphs and the travelling salesman problem", *Proceedings of the American Mathematical Society* 7, 48–50.

Labbé, M., Peeters, D., and Thisse, J.-F. (1995), "Location on networks", in: M. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, eds., *Handbooks in Operations Research and Management Science 8. Network Routing*, North-Holland, Amsterdam, 551–624.

Laporte, G. (1992a), "The traveling salesman problem: An overview of exact and approximate algorithms", *European Journal of Operational Research* 59, 231–247.

Laporte, G. (1992b), "The vehicle routing problem: An overview of exact and approximate algorithms", *European Journal of Operational Research* 59, 345–358.

Laporte, G., and Louveaux, F.V. (1993), "The integer L-shaped method for stochastic integer programs with complete recourse", *Operations Research Letters* 13, 133–142.

Laporte, G., and Nobert, Y. (1987), "Exact algorithms for the vehicle routing problem", *Annals of Discrete Mathematics* 31, 147-184.

Laporte, G., Nobert, Y., and Desrochers, M. (1985), "Optimal routing under capacity and distance restrictions", *Operations Research* 33, 1050-1073.

Magnanti, T.L., and Wolsey, L.A. (1995), "Optimal trees", in: M. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, eds., *Handbooks in Operations Research and Management Science 6. Network Models*, North-Holland, Amsterdam, 503-615.

Magnanti, T.L., and Wong, R.T. (1986), "Network design and transportation planning: Models and algorithms", *Transportation Science* 18/1, 1-55.

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

Nagurney, A. (1993), *Network Economics - A Variational Inequality Approach*, Kluwer, Dordrecht.

Nelson, M.D., Nygard, K.E., Griffin, J.H., and Shreve, W.E. (1985), "Implementation techniques for the vehicle routing problem", *Computers and Operations Research* 12/3, 273-283.

Osman, I.H., and Kelly, J.P., eds. (1993), *Meta-Heuristics: Theory & Applications*, Blackwell Scientific Publications, Oxford.

Owoc, M., and Sargious M.A. (1992), "The role of transportation in free trade competition", in: N. Waters, ed., *Canadian Transportation: Competing in a Global Context*, 23-32.

Pirlot, M. (1992), "General local search heuristics in combinatorial optimization. A tutorial", *Belgian Journal of Operations Research, Statistics and Computer Science* 32/2, 8-67.

Powell, W.B. (1988), "A comparative review of alternative algorithms for the dynamic vehicle allocation problem", in: B.L. Golden and A.A. Assad, eds., *Vehicle Routing: Methods and Studies*, North-Holland, Amsterdam, 249-292.

Powell, W.B. (1995), "Toward a unified modeling framework for real-time control of logistics", *Military Journal of Operations Research*, forthcoming.

Powell, W.B. (1996), "A stochastic formulation of the dynamic assignment problem, with an application to truckload motor carriers", *Transportation Science* 30/3, 195-219.

Powell, W.B., and Cheung, R.K.-M. (1994a), "A network recourse decomposition method for dynamic networks with random arc capacities", *Networks* 24, 369-384.

Powell, W.B., and Cheung, R.K.-M. (1994b), "Stochastic programs over trees with random arc capacities", *Networks* 24, 161-175.

Powell, W.B., and Cheung, R.K.-M. (1996), "An algorithm for multistage dynamic networks with random arc capacities, with an application to dynamic fleet management", *Operations Research* 44/6.

Powell, W.B., and Frantzeskakis, L. (1994), "Restricted recourse strategies for dynamic networks with random arc capacities", *Transportation Science* 28/1, 3-23.

Powell, W.B., and Sheffi, Y. (1989), "Design and implementation of an interactive optimization system for the network design in the motor carrier industry", *Operations Research* 37/1, 12-29.

Powell, W.B., Sheffi, Y., Nickerson, K.S., Butterbaugh, K., and Atherton, S. (1992), "Maximizing profits for North American Van Lines' Truckload Division: A new framework for pricing and operations", *Interfaces* 18/1, 21-41.

Powell, W.B., Carvalho, T.A., Godfrey, G.A., and Simaõ, H.P. (1994), "Optimal control methods for logistics queueing networks", in: *TRISTAN II, Triennial Symposium on Transportation Analysis*, Preprints, 185-199.

Powell, W.B., Carvalho, T.A., Godfrey, G.A., and Simaõ, H.P. (1995a), "Dynamic fleet management as a logistics queueing network", *Annals of Operations Research* 61, 165-188.

Powell, W.B., Jaillet, P., and Odoni, A. (1995b), "Stochastic and dynamic networks and routing", in: M. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, eds., *Handbooks in Operations Research and Management Science 8. Network Routing*, North-Holland, Amsterdam, 141-295.

Psaraftis, H.N. (1995), "Dynamic vehicle routing: Status and prospects", *Annals of Operations Research* 61, 143-164.

Quinet, E. (1990), *Analyse économique des transports*, Presses universitaires de France, Paris.

Reeves, C.R., ed. (1996), *Modern Heuristic Techniques for Combinatorial Problems*, Kluwer Academic Publishers, Norwell, MA.

Rochat, Y., and Taillard, E. (1995), "Probabilistic diversification and intensification in local search for vehicle routing", *Journal of Heuristics* 1/1, 147-167.

Roy, J. (1984), "Un modèle de plannification globale pour le transport routier des marchandises", Ph.D. Thesis, Ecole des Hautes Etudes Commerciales, Université de Montréal, Montréal, Canada.

Roy, J., and Crainic, T.G. (1992), "Improving intercity freight routing with a tactical planning model", *Interfaces* 22/3, 31-44.

Roy, J., and Delorme, L. (1989), "NETPLAN: A network optimization model for tactical planning in the less-than-truckload motor-carrier industry", *INFOR* 27/1, 22-35.

Sheffi, Y. (1985), *Urban Transportation Networks. Equilibrium Analysis with Mathematical Programming*, Prentice-Hall, Englewood Cliffs, NJ.

Shen, S., Potvin, J.-Y., Rousseau, J.-M., and Roy, S. (1995), "A computer assistant for vehicle dispatching with learning capabilities", *Annals of Operations Research* 61, 189-211.

Taff, C.A. (1978), *Management of Physical Distribution and Transportation*, Irwin, Homewood, IL.

Taillard, E. (1993), "Parallel iterative search methods for vehicle routing problems", *Networks* 23, 661-673.

Turnquist, M.A. (1994), "Economies of scale and network optimization for empty car distribution", Presented at the TIMS/ORSA National Meeting, Detroit, MI.

Yellow, P. (1970), "A combinatorial modification to the savings method of vehicle scheduling", *Operational Research Quarterly* 21, 281-283.

Zalatan, P. (1993), "Economic cycles, structural change and the transportation sector", in: B.G. Bison, ed., *A Look Back from the Year 2000*, 28th Annual Meeting of the Canadian Transportation Research Forum, 111-121.