Transpn. Res.-B Vol. 20B, No. 3, pp. 225–242, 1986
Printed in Great Britain.

0191-2615/86 $3.00+.00

MULTICOMMODITY, MULTIMODE FREIGHT TRANSPORTATION: A GENERAL MODELING AND ALGORITHMIC FRAMEWORK FOR THE SERVICE NETWORK DESIGN PROBLEM

TEODOR G. CRAINIC and JEAN–MARC ROUSSEAU
Centre de Recherche sur les Transports, Université de Montréal, P.O. Box 6128, Station A, Montréal, Québec Canada H3C 3J7

(*Received* 20 *July* 1984; *in revised form* 31 *January* 1985)

Abstract—We examine the freight transportation problem which occurs when the same authority controls and plans both the supply of transportation services (modes, routes, frequencies for the services and classification, consolidation, transfer policies for terminals) and the routing of freight. We present a general modeling framework, based on a network optimization model, which may be used to assist and enhance the tactical and strategic planning process for such a system. The problem is solved by means of an algorithm, described in some detail, based on decomposition and column generation principles. We also present detailed results on the behaviour and performance of the algorithm, as observed during experimentation with a specific rail application.

1. INTRODUCTION

In this paper, we examine the multimode, multicommodity freight transportation problem which occurs when the same authority supplies or regulates the supply of transportation services (including terminal operations) and also controls, at least partially, the routing of the goods through this service network.

We intend our model to be used at the medium-term (or tactical) planning level and we are interested in the major problems within the scope of decision making at this level: the design of the service network, the development of terminal policies and the establishment of traffic routing patterns through the network.

Our aim is to design a global model which integrates these problems, describes their interrelationships and network-wide consequences and considers the tradeoffs to be made between economic and service (time) criteria.

The objective is the enhancement of the planning process and hence the improvement of the performance of the transportation system, by making possible the efficient and systematic generation and evaluation of a large number of global operating strategies. The purpose of the model, therefore, is not to obtain a detailed representation of operations but to generate "best" operating strategies to reduce costs and delays and to improve the quality of service.

Among the possible applications of such a model, the following areas can be suggested.

(i) Freight transportation by rail: here, the different types of train service (normal, rapid, direct, unit, etc.) correspond to the various "modes";
(ii) Intercity transportation by truck: this is the "less-than-truckload" problem with or without multitrailer convoys and with or without TOFC;
(iii) freight transportation in developing countries where a central authority controls, more or less, the entire transportation system;
(iv) TOFC/COFC (trailer-on-flatcar/container-on-flatcar) intermodal transportation;
(v) container transportation through a combination of air/sea/highway/rail modes.

The paper begins with a presentation of the general problem and a brief literature review. The modeling and algorithmic framework are presented in the next section, while the following sections detail the different procedures that make up the algorithm, as well as some of the heuristic strategies we have adopted to improve its performance. We then present the specific rail application which prompted our work, together with some numerical results and performance measures obtained for this problem. The paper ends with some general remarks and comments.

2. THE PROBLEM

As usually defined, tactical planning aims at the efficient and rational allocation of existing resources, over a medium-term planning horizon, to improve the overall performance of a system.

In freight transportation, tactical planning focuses mainly on the following major problems:

(i) Service network design or carrier (service) routing is concerned with the type and level of service to be offered. On what routes should it be offered, of what type (mode) and how often?

(ii) Traffic (freight) routing determines how the traffic moves through the network. For each traffic-class (an origin-destination-commodity combination), the route or routes through the service network, the terminals on these routes and the amount of freight using each route must be determined.

(iii) Terminal policies determine the strategies for the classification (or consolidation) of traffic, for the modal transfer and for the assignment of traffic to carriers at each terminal of the system.

These problems and the policies addressing them have network-wide impacts and are strongly and complexly interconnected both in economic terms and in their space-time dimensions. So, to fully capture the effects of these policies on the global performance of the transportation system, both operating costs and some measure of the quality of service (usually, the delays imposed on freight and carriers) have to be included in the decision-making process. Moreover, tradeoffs between the operating costs implied by a certain policy and its impact on the level and quality of the service have also to be considered. The same tradeoffs are also important when the relationships between policies addressing the three main problems considered in tactical planning are investigated.

The modeling framework proposed is thus directed towards capturing these interactions and reflecting these tradeoffs. It takes the form of a network optimization model aimed at generating best global strategies to reduce operating costs and delays and improve global system performance.

There are few papers to be found in the literature that directly address these same issues. Recent reviews of freight transportation related work may be found in Friesz *et al.* (1983) (for a review of predictive intercity freight models) and Crainic *et al.* (1984b) (for a review of methodologies used in freight transportation studies).

The technological aspects of intermodal freight transportation are the subject of a large number of papers and a review of existing and developing technology, together with an annotated bibliography, may be found in Vecellio *et al.* (1981).

Work on intermodal, multicommodity freight transportation has mainly been conducted in the context of regional or national planning. In most of these studies however, the emphasis is either on the routing of traffic (and, eventually, the modal split) through the exogenously fixed multimodal service network (Kresge and Roberts, 1971; McGinnis *et al.*, 1981), or on the investigation of the equilibrium relations among the supply of transportation services and the shippers' demand for these services in a competitive environment: the Transportation Network Model (CACI Inc., 1980; Bronzini, 1980), the Freight Network Equilibrium Model and its extentions (Friesz *et al.* 1981; Harker, 1983; Friesz *et al.*, 1985) and the recent work of Fisk and Boyce (1983). In both cases, the problem addressed is different from the one examined in this paper. We believe, however, that our method may be advantageously adapted to address the first type of problems.

Problems somewhat similar to ours may be found in the context of passenger transportation by train, bus or airplane (see, for example, Soumis *et al.*, 1980 and 1981; Powell, 1982) where, however, the assumptions concerning the passenger's behavior are quite different from those applicable to the movement of goods in our problem. Moreover, terminal operations in freight transportation are both more complex and more frequent than in passenger transportation.

The less-than-truckload aspect of freight transportation by truck is one of the problems which may be addressed by our method. Powell and Sheffi (1983) and Roy (1984) have recently made significant contributions to this field. Powell and Sheffi study the problem of establishing the load plan, that is, the routing of the shipments through a sequence of consolidation terminals. The problem is formulated as a very large mixed integer programming model and is solved by a heuristic based on a local improvement approach. Roy (1984) studies the more general tactical planning problem using the modeling and algorithmic framework described in this paper.

Our research was initially prompted by a rail freight transportation problem for a large Canadian company. In the context of rail transportation, the review by Assad (1980a) should be mentioned. Important work related to the tactical planning of rail freight transportation has been done by Thomet (1971a, b), Petersen *et al.*, (1975), Bodin *et al.*, (1980) and Assad (1980b).

These writers have all made significant contributions to the field of network rail modeling. We believe our approach to be more comprehensive, however, as it can address a larger class of problems. In Crainic *et al.*, (1984a) we presented the detailed model for the rail case together with the scheme of the algorithm. In the following pages, we present a generalization of this model. We also detail and analyze the algorithm and its performance.

3. THE MODELING AND ALGORITHMIC FRAMEWORK

Consider a physical network $G_{\text{ph}} = (N, A_{\text{ph}})$ where N is the set of nodes (mode-dedicated or intermodal terminals, pick-up/delivery points, junctions), and A_{ph} is the set of physical (road segments, rail tracks, river segments) or conceptual (possible air or sea lines) links joining these nodes.

On the physical network, the demand for transportation is assumed to be specified for each traffic class in terms of a certain measure of tonnage of freight (or number of vehicles), of the specified commodity type, to be moved from the origin terminal to the destination terminal.

The service network $G_s = (N, A_s)$ specifies the transportation services that might be offered to satisfy the demand for transportation. Each service, $a_h \in A_s$, is defined by the route it follows through the physical network from its origin to its destination, by the sequence of terminals where it stops on this route and by its type of service (specifying characteristics such as mode, speed, priority, etc.). The other important service characteristic is the "frequency" which defines the level of service that is to be offered on that route during the planning period. The values of the frequencies will be determined by the model and a zero frequency will imply that the corresponding service will not be offered. Figure 1 illustrates some of the feasible services on a simple, linear physical network.

An itinerary defines a feasible journey for the freight of a given traffic class: the sequence of terminals (composed of the origin, destination and, possibly, some intermediate terminals) where certain operations (classification, consolidation, etc.) are to be performed on the traffic, plus the paths of carrier services in G_s used to move the freight between two consecutive terminals in this sequence. Some examples are illustrated in Fig. 1. The amount of flow to be assigned to every such itinerary will also be determined by the model. The number of possible itineraries for each traffic class may be extremely large. Therefore, a column generation procedure has been included in the algorithm to ensure that only a limited number of good itineraries are generated, as and when required during the resolution. This procedure is described in Section 5.

A formal model may then be written as follows. Minimize

$$\Psi(x, t) = \sum_m \sum_k Z_k^m(x, t) + \sum_h Y_h(t) + P(x, t)$$

subject to

$$\sum_k x_k^m = d^m, \qquad \text{all } m$$

$$x_m \geqslant 0, \qquad \text{all } m$$

$$t_h \geqslant 0 \text{ and integer}, \qquad \text{all } h,$$

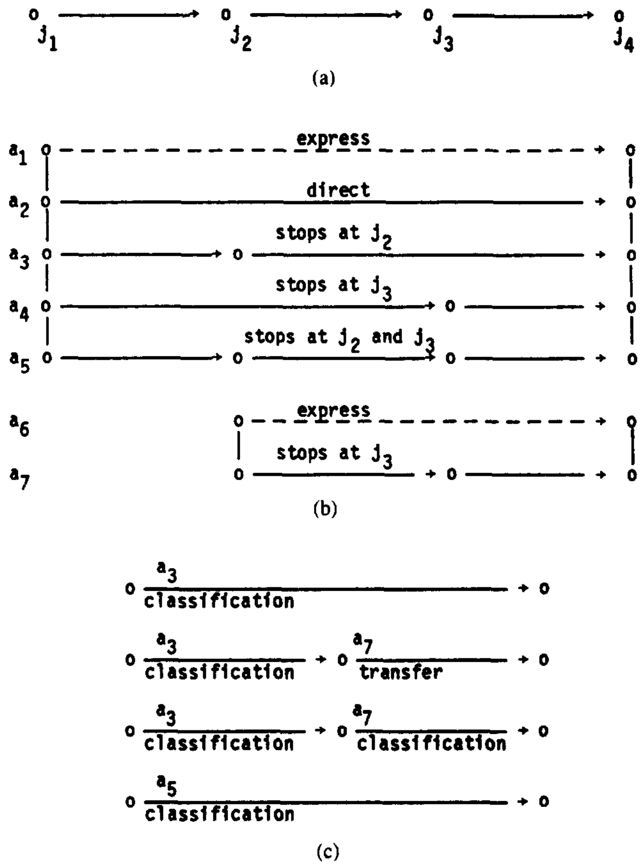

Fig. 1. Examples of services and itineraries. (a) $G_{ph} = (N, A_{ph})$. (b) Some services of $G_s = (N, A_s)$. (c) Four itineraries for the traffic class (j_1, j_3, c).

where the indices h, m and k refer, respectively, to services, traffic classes and itineraries for each traffic class, while

x_k^m = quantity of flow from the traffic class m traveling on itinerary k,
d^m = demand for transportation for the traffic class m,
t_h = frequency of the service a_h,
x_m = vector of flows for the traffic class m,
x = vector of flows for the whole system (by traffic class),
t = vector of frequencies,
$Z_k^m(x, t)$ = total (operating and delay) cost of using the itinerary k for the traffic-class m; the sum represents the "variable" cost of transporting the flow x by a service network operated at level t,
$Y_h(t)$ = total (operation and delay) cost of operating the service a_h; the sum represents the "fixed" cost of offering the transportation services at the frequencies t,
$P(x, t)$ = special terms modeling specific relations, such as train capacity, as delay costs.

Thus, the objective function mainly represents the total system cost: traffic and itinerary related costs, plus the cost of offering the transportation service. It is a generalized cost, in the sense that both operating and delay costs are included. Delay terms are transformed into delay costs compatible with operating costs, via user-defined unit time costs for each traffic class, type of service, type of vehicle. These unit costs, generally based on equipment depreciation values and goods inventory costs, may also be used to reflect time-related characteristics of the traffic, such as priorities among traffic classes or time sensitiveness for certain traffic classes.

Among the terms that are most likely to appear in the objective function, are the following.

(i) Delays in terminals generated by the various operations (and the resulting congestion) on carriers and traffic: traffic classification, carrier loading and possibly, convoy formation, interservice traffic transfer. Frequency (or connection) delays also have to be included. They represent the time spent waiting for the designated service to be available and are often used as a measure of the quality of the service offered.

(ii) Over-the-road (or longhauling) delays affecting both traffic and carriers and representing the interterminal transportation time plus possible delays due to congestion on the physical network.

(iii) Operating costs (manpower, energy, etc.) due to traffic and carrier handling in terminals and on the links.

The objective function also includes terms, which we have designated by $P(x, t)$, modeling additional relations between the elements of the model. A prime example is given by the capacity relations between the services and the quantity of freight they transport. These relationships are generally formulated as strict capacity constraints, especially when modeling operational problems such as the routing and scheduling of delivery vehicles. For a tactical planning model however, one is generally less concerned with the specific vehicle capacity, the emphasis rather being on determining the frequency of the service, which defines the service capacity, and the traffic flow distribution, which determines how this capacity is utilized. Thus, the assignment of some traffic to a particular service resulting in an exceeding of the service capacity may indicate either that the frequency of the service should be increased, or that some less important (in terms of priority or delay costs) traffic should pass to another service. The formulation of strict capacity constraints would prevent, however, an easy detection and manipulation of such a situation.

It is thus more natural, in this case, to model these relations as utilization targets where overcapacity assignment is permitted at the expense of additional costs and delays. We have adopted this latter view and modeled these relations as additional delays to the traffic, computed by penalty-like terms on the quantity of flow assigned over and above the predesignated (service and possibly link and commodity specific) capacity of the service. Other relations, such as fleet availability constraints or minimum service performance targets, may also be treated in a similar way.

The particular functional forms of the terms of the objective function are application specific and will not be discussed here. We do require, however, that, on the intervals of interest, the functions present some differentiability characteristics required by the algorithmic techniques presented later on in this paper. Examples of detailed formulations following this modeling framework may be found in Crainic *et al.* (1984a) and Roy (1984).

We now want to briefly address the backhauling, or empty vehicle movement, problem. It is a particular problem for organizations offering freight transportation services, which arises from imbalances in the flow of goods, generating the need to move empty vehicles (rail cars, trucks, trailers, containers, etc.) to satisfy the demand. One way of considering this problem in a tactical planning model is first to compute an empty vehicle demand matrix and then to consider empties as a commodity in the tactical model. The computation of this matrix may be done by a two-step procedure: (i) compute, from the O/D matrix, the supply and demand of vehicles (possibly, by type) at each node of the network; (ii) find the O/D distribution of empties by, for example, a gravity model (see Coutu, 1978). This procedure is adequate for the rail and container problems and is used, for example, at Canadian National Railways. For the truck problem, where a more complex approach is needed, Roy (1984) presented a procedure where the model is expanded to include the cost of empty returns and the equilibration of the number of arrivals and departures, at each particular terminal.

Finally, two important characteristics of the formulation should be stressed:

(i) The explicit and simultaneous consideration of the two levels of planning—the design of the service network (and the determination of the service level) and the traffic distribution problem.

(ii) The modularity of the formulation (matched by the modularity of the algorithm) which allows flexibility in the definition of the service network and of the feasible traffic routing as well as, in the choice of the particular terms (activities, delays, operating costs) to be included in the objective function and the specific functional forms for these terms.

It is also important to note that this formulation is not only a tactical model but may also be used as an evaluation model for scenarios defined during strategic planning and involving modifications in the physical network, major changes in policies and regulations, etc.

The formulation yields a nonlinear, mixed integer, multicommodity flow problem with a special structure: explicit constraints bind only the continuous variables with the resulting constraint matrix having a block-diagonal form. Each block corresponds to a given traffic class and has only one row.

Consequently, we solve this problem via a decomposition-based algorithm which works alternately on the following problems.

1. Service network design: given a fixed traffic distribution, modify the frequencies to improve the objective function, if possible. If this is not possible the algorithm stops.

2. Traffic routing: given a certain level of service, in terms of the frequencies offered, determine the best traffic distribution.

We now proceed with a description of the major procedures which compose this algorithm.

4. MODIFYING THE SERVICE NETWORK

The service network design problem is our master problem, in the sense that it contains the termination criterion and controls the feasibility region for the traffic distribution problem by determining which routes are available and the service level offered.

Form of the algorithm

The algorithm may therefore be written in the following general form:

1. Let t^0 be an initial service level and x^0 the optimum (in the sense of the optimization described in section 5) traffic assignment on the service network.

2. At iteration k (i) fix the traffic flow variables at their current values,

$$\bar{x} = x^{k-1};$$

(ii) find the new level of service,

$$t^k \leftarrow p(t^{k-1}, \bar{x}),$$

such that

$$\Psi(\bar{x}, t^k) \leq \Psi(\bar{x}, t^{k-1});$$

(iii) if this is not possible, stop; (iv) otherwise find x^k, the optimal traffic assignment for the service level t^k.

The traffic distribution procedure, which computes the traffic flows x^k, is described in section 5. The new level of service t^k is found by a heuristic procedure called the frequency modification procedure and described in the next section. This heuristic also gives a lower bound on the overall improvement in the value of the objective function and ensures that the integrality conditions on the frequency variables are satisfied.

Frequency modification procedure

Consider the vector of finite differences $\Pi(\Delta)$, where Δ is a positive integer, s is the total number of services and

$$\begin{aligned}\Pi_h(\Delta) = {} & \Psi(\bar{x}, t_1, \ldots, t_{h-1}, t_h, t_{h+1}, \ldots, t_s) \\ & - \Psi(\bar{x}, t_1, \ldots, t_{h-1}, t_h - \Delta, t_{h+1}, \ldots, t_s) \qquad \text{for all } h.\end{aligned}$$

Then, $\Pi_h(1)$ represents the marginal value of decreasing the frequency of the service a_h by one unit, with no change in the traffic pattern. If this marginal value is positive, it also represents a lower bound on the improvement to be obtained in the value of the objective function, should that modification be implemented. For, suppose we have a solution (x^k, t^k) and that, for a certain $h \epsilon [1, s]$, there is a $\Delta > 0$ such that $\Pi_h(\Delta) > 0$. Define the new vector of frequencies

$$t_j^{k+1} = \begin{cases} t_j^k, & \text{if } j \neq h \\ t_j^k - \Delta, & \text{if } j = h \end{cases}.$$

Then, evidently,

$$\Psi(\bar{x}, t^{k+1}) \leqslant \Psi(\bar{x}, t^k).$$

Also, suppose that x^{k+1} is the optimal traffic assignment for the service level t^{k+1}. Then, remembering that x^k is an optimal traffic assignment for t^k and that the traffic assignment problem minimizes the total cost, it can be seen that modifying the frequencies from t^k to t^{k+1} can result in only one of two mutually exclusive alternatives:

(i) x^k is still an optimal traffic assignment for the new service level t^{k+1}; then $x^k = x^{k+1}$ and $\Psi(x^{k+1}, t^{k+1}), = \Psi(\bar{x}, t^{k+1})$.

(ii) x^k is different from x^{k+1}; this means that the modification performed on the level of service has allowed a better traffic assignment to be found and, consequently, a better overall solution, i.e.,

$$\Psi(x^{k+1}, t^{k+1}) < \Psi(\bar{x}, t^{k+1}).$$

Then,

$$\begin{aligned}\Pi_h(\Delta) &= \Psi(\bar{x}, t^k) - \Psi(\bar{x}, t^{k+1}) \\ &\leqslant \Psi(x^k, t^k) - \Psi(x^{k+1}, t^{k+1}).\end{aligned}$$

Consequently, the heuristic approach we propose initializes all frequencies at relatively high values (the idea being that all services should be available at the beginning of the process) and gradually decreases them according to the impact of their decrease on the value of the objective function as estimated by $\Pi(\Delta)$. Then, at each iteration, the new vector of frequencies is obtained by the following basic frequency modification procedure:

(i) Compute $\Pi(1)$;
(ii) Find $u_h = \underset{1 \leqslant k \leqslant s}{\text{maximum}} \{\Pi_k(1) | \Pi_k(1) > 0\}$;
(iii) If $u_h \leqslant \xi$, for a given, small $\xi \geqslant 0$, then stop.
(iv) If not, modify the frequencies

$$t_j = \begin{cases} t_j, & \text{if } j \neq h \\ t_j - \Delta, & \text{if } j = h \end{cases}$$

where Δ is a predetermined positive integer. The determination of the step size Δ is a heuristic process which is explained through a discussion of the behaviour of $\Pi_h(\Delta)$.

First, when the frequency of a given service is decreased, the cost of operating that service is also, obviously, decreased. At the same time, the congestion on all the links making up its route is decreased, so the over-the-road delays for all services and traffic traveling on these links will also decrease (suppose, for the moment, that the capacity-penalty terms for that service are not active, nor are they activated by the modification). On the other hand, a decreased level of service will also generate longer delays at terminals (frequency delays) for all traffic traveling on that particular service. Generally speaking, however, it is the decrease in over-the-road delays and, especially, in carrier operating costs which dominate. This is particularly true under "heavy traffic" conditions such as those created by our initialization procedure.

Furthermore, considering that carrier operating costs are generally taken to be linear functions of the level of service, while over-the-road delays are approximated by monotone increasing congestion functions, the finite differences $\Pi_h(\Delta)$ also exhibit a monotone behaviour [providing $\Pi_h(1)$ is strictly positive] on some interval $[\Delta_{min}, \Delta_{max}]$. Here, Δ_{min} is negative and represents the maximum frequency increment before jamming occurs (i.e., there are so many vehicles on the route that the over-the-road delays are infinite for all, or at least some, types of services) or the maximum service level is attained (if any has been defined). On the other hand, Δ_{max} is positive and represents the maximum frequency decrement before the capacity penalty is activated or the increase in the frequency delay costs for traffic offsets the other improvements. In fact, during our tests, linear approximations of $\Pi(h)$ (obtained from regression models) for different services, under various congestion conditions, generated errors (as measured by the mean square error of the regression equation) of the order of 10^{-4}, which seems to indicate an almost linear behaviour.

Thus, depending upon the stage reached in the solution process, different values for Δ may be used: larger values (such as half the value of Δ_{max}) are used in the early stages when Δ_{max} is relatively important, while a unit-valued Δ is used when Δ_{max} is small.

At this point, we stress again that each $\Pi_h(\Delta)$ is computed with the values of all other variables, frequencies and traffic flows, being kept fixed. Consequently, following the frequency modification of a certain service a_h for which $\Pi_h(\Delta) > 0$, a redistribution of flows generally does improve the solution and thus $\Pi_h(\Delta)$ underestimates the total improvement due to the frequency modification $t_h = t_h - \Delta$. Note that, by the same reasoning, the objective function also may be improved after performing $t_h = t_h - \Delta$ even when $\Pi_h(\Delta) < 0$: a redistribution of flows might bring an overall improvement larger than $\Pi_h(\Delta)$. This proved indeed to be the case when, toward the end of the solution process, all $\Pi_h(\Delta)$ were nonpositive (for positive and negative Δ), while some services with positive frequencies transported only residual traffic. By removing these services and forcing the rerouting of the corresponding traffic, some improvement was obtained. However, this phenomenon was not observed as long as there was some positive $\Pi_h(\Delta)$ indicating a possible improvement.

We now briefly examine the problem of increasing the frequency of a service a_h. In this case, the finite difference $\Pi_h(-1)$ still represents the marginal value of increasing the frequency t_h by one unit while all other frequencies and the traffic distribution pattern are fixed at their current values. In general, however, $\Pi_h(-1)$ is not a good estimation of the expected improvement in the value of the objective function, due to the increase of the frequency t_h by one unit.

This is because, by the same arguments, *mutatis mutandis*, developed for the $\Pi_h(1)$ case, increasing a frequency will often produce, especially in heavy traffic conditions, a negative value for the finite difference $\Pi_h(-1)$.

Furthermore, the attraction phenomenon may come into play: increasing the frequency of a service might make it attractive for traffic traveling currently on other services. Changes in the traffic distribution might thus occur, followed by frequency readjustments for the services which have lost traffic, with the net result of a decrease in the value of the objective function.

It is thus possible to improve the objective by increasing a frequency even though $\Pi_h(-1)$ is negative. This involves a quite complicated process however, whose benefits are very poorly reflected by the behaviour of $\Pi(-\Delta)$.

We have thus decided to mainly follow the strategy of gradually decreasing the frequencies from high initial values, strategy which facilitates the estimation, via the finite differences $\Pi(\Delta)$, of the expected benefits of modifications of frequencies.

Yet, under "light traffic" conditions, such as those in effect in the late stages of the solution process or when the freight delay costs are high enough, the value of $\Pi(-\Delta)$ could be meaningful. At that moment, we therefore use the following variant of the basic frequency modification procedure:

(i) compute $\Pi(1)$ and $\Pi(-1)$;
(ii) find

$$\begin{aligned} u_h &= \underset{1\leq j\leq s}{\text{maximum}} \{\Pi_j(1)|\Pi_j(1) \geq 0\}, \\ u_k &= \underset{1\leq j\leq s}{\text{maximum}} \{\Pi_j(-1)|\Pi_j(-1) \geq 0\}, \\ u &= \text{maximum} \{u_h, u_k\}; \end{aligned}$$

(iii) if $u \leq \xi$, for a given small $\xi \geq 0$, then stop;
(iv) if not, modify the frequencies

$$t_j = \begin{cases} t_j - 1, & \text{if } u = u_h, j = h \\ t_j + 1, & \text{if } u = u_k, j = k. \\ t_j, & \text{otherwise} \end{cases}$$

Naturally, more than one service may have its frequency modified each time the frequency modification procedure is executed, provided the services affected do not share any route link. Of course, the greater the number of services handled at the same time, the greater the time and effort spent verifying their disjointness. This time and effort is related to the size and intricacy of the service network, to the topology of the physical network (linear or with many parallel routes) and to the computer memory space available and compromises must be reached. In our case, we have adopted a procedure which modifies only two services at a time: the best (i.e., the service whose finite difference is maximum for the chosen step size) and the second disjoint best.

5. OPTIMIZING THE TRAFFIC DISTRIBUTION

The second main component of the algorithm reoptimizes the routing of the freight, once one or more frequencies have been modified. The frequencies are then fixed at their current values $\bar{t}$ and the problem to be solved reduces to the following: minimize

$$\Psi(x, \bar{t})$$

subject to

$$\begin{aligned} \sum_k x_k^m &= d^m, && \text{for all } m \\ x_k^m &\geq 0, && \text{for all } m, \text{ all } k. \end{aligned}$$

Taking advantage of the structure displayed by the formulation, we solve this problem by a decomposition procedure which includes an itinerary generation algorithm.

Decomposition procedure

1. Cyclically, for $m = 1, 2, \ldots, M$

(i) fix the other flow variables at their current values $\bar{x}_s$, $s \neq m$;
(ii) generate i_m, the "best" itinerary for the traffic class m, given the present conditions;

(iii) if $\{i_m\} \cap I_m = \emptyset$ then $I_m = I_m \cup \{i_m\}$;
(iv) solve the new subproblem (*): minimize

$$\phi(X_m),$$

subject to

$$\sum_h x_k^m = d^m, \; x_k^m \geq 0 \qquad \text{for all } k \text{ in } I_m.$$

2. If, for all m, there is a v_m such that, for all k in I_m,

$$\left.\frac{\partial \Psi(x, \bar{t})}{\partial x_k^m}\right|_{x_k^m = \bar{x}_k^m} \begin{cases} = -v_m, & \text{if } \bar{x}_k^m > 0 \\ \geq -v_m, & \text{if } \bar{x}_k^m = 0 \end{cases}$$

then stop; otherwise, repeat the first step. Where

M = total number of traffic-classes,
I_m = the set of itineraries for the traffic class m,
$\phi(x_m) = \Psi(\bar{x}_1, \ldots, \bar{x}_{m-1}, x_m, \bar{x}_{m+1}, \ldots, \bar{x}_M, \bar{t})$.

A solution $\bar{x}$ satisfying the termination criterion of this procedure is a feasible Kuhn–Tucker point and is called a stable solution. Note that this condition represents the union of the equivalent conditions for the M individual subproblems. Thus, the procedure will stop when, in the course of the same cycle, a stable solution condition has been satisfied for all traffic classes.

From a practical point of view, the procedure may be stopped earlier, as soon as the conditions are almost satisfied. The following criterion may be used then: Stop when, for a prespecified small $\xi \geq 0$ and for all m,

$$\left| \underset{x_k^m > 0}{\text{maximum}} \left\{ \frac{\partial \phi(x_m)}{\partial x_k^m} \right\} - \underset{x_k^m > 0}{\text{minimum}} \left\{ \frac{\partial \phi(x_m)}{\partial x_k^m} \right\} \right| < \xi,$$

$$\underset{x_k^m = 0}{\text{minimum}} \left\{ \frac{\partial \phi(x_m)}{\partial x_k^m} \right\} - \underset{x_k^m > 0}{\text{maximum}} \left\{ \frac{\partial \phi(x_m)}{\partial x_k^m} \right\} > \xi.$$

To solve the subproblem (*), descent methods (see Bazarra and Shetty, 1977) seem appropriate. Owing to the particular structure of the problem, analytical expressions may be relatively easily derived for the descent directions and an efficient adaptation of the gradient method may be implemented (see Crainic, 1982).

Finally, note that should $\Psi(x, \bar{t})$ be a (strictly) convex function, our procedure would find the (unique) optimal traffic assignment, given the present level of service and the itineraries already generated. Unfortunately, this is generally not the case and the decomposition procedure is therefore a heuristic converging only to a local solution. In our tests, however, the results indicate that the functions are reasonably "well behaved" and that good solutions are obtained.

Itinerary generation process

The objective of the itinerary generation algorithm is to obtain new variables that might improve the current solution. Thus, we look for the best itinerary for the traffic class under consideration, given the current level of service and the present congestion conditions on the network.

By our definition, the best itinerary offers the lowest possible total cost (i.e., operating plus delay cost) per unit of freight flow. Note that this is, in fact, a marginal cost, in the sense that it is the lowest total cost for the first unit of flow that will be assigned to the new itinerary.

Also by definition, an itinerary is specified by a sequence of working (classification) terminals and the path of services between any two consecutive terminals in that sequence.

For the column generation process we therefore propose a shortest-path-like algorithm which uses as average unit cost for each facility (terminal or service link) the cost of the last unit of flow which used that particular facility.

The algorithm works on a conceptual network $G_l = (N, A_l)$, where N is the same node set as before, while the set of arcs A_l is such that an arc $(i, j) \in A_l$, $i, j, \epsilon N$, only if there is a path in the service network G_s from terminal i to terminal j.

On this network, $c(i, j)$, the cost associated with a unit of flow traveling on arc $(i, j) \in A_l$, is defined as the cost of $ch(i, j)$, the path with the lowest total cost from terminal i to terminal j, in the service network G_s.

In addition, a unit of traffic which passes through and is classified at terminal j incurs a cost (as always, we include both operating and delay costs) which may be decomposed into two components: (i) a fixed cost which depends exclusively on current conditions at terminal j (congestion, physical and technological characteristics, etc.), and (ii) a variable cost which is a function of the destination terminal for which the traffic is classified.

We then define $S_j = \{\beta(j) | \beta(j) \epsilon N$ is an admissible classification destination for the traffic classified at node j, $j \in N\}$; $\gamma_j(\beta(j))$: unitary cost of classifying at terminal j for terminal $\beta(j)$, j, $\beta(j) \in N$.

Then the best itinerary from a terminal P to another terminal Q may be found by the following labeling procedure:

1. Label each node $j \in N$,

$$[\alpha(j), \tau(j)],$$

where $\alpha(j)$ = predecessor of j on the best itinerary from P to j; $\tau(j)$ = the cost of the best itinerary from P to j;

$$\alpha(j) = -\text{for all } j \in N;$$

$$\tau(j) = \begin{cases} 0, & \text{if } j \equiv P \\ \infty, & \text{otherwise} \end{cases};$$

$S_Q \equiv \emptyset$; P is a marked node and all others are unmarked;

2. Suppose $j_0 \equiv P, j_1, \ldots, j_{k-1}$ are previously marked nodes. For each marked node j_i, $0 \leq i \leq k - 1$, choose a *candidate node* $\beta^*(j_i) \in S_{j_i}$ and $\beta^*(j_i)$ unmarked, such that

$$\gamma_{j_i}(\beta^*(j_i)) + c(j_i, \beta(j_i)) = \underset{\substack{\beta(j_i) \in s_{j_i} \\ \text{unmarked}}}{\text{minimum}} \{\gamma_{j_i}(\beta(j_i)) + c(j_i, \beta(j_i))\}$$

Identify the node $j_k = \beta^*(j_l)$, $0 \leq l \leq k - 1$ such that

$$\tau(j_l) + \gamma_{j_l}(\beta^*(j_l)) + c(j_l, \beta^*(j_l)) = \underset{0 \leq i \leq k-1}{\text{minimum}} \{\tau(j_i) \; ; \gamma_{j_i}(\beta^*(j_i)) + c(j_i, l^* \; (j_i)\beta^*(j_i))\}.$$

Mark the node j_k and label it:

$$\alpha(j_k) = j_l,$$

$$\tau(j_k) = \tau(j_l) + \gamma_{j_l}(j_k) + c(j_l, j_k).$$

3. Repeat until Q is marked and labeled.

If there are no negative cycles in G_l and if $\tau(Q) \neq \infty$, a simple inductive argument shows that the best itinerary from P to Q is

$$j_o, ch(j_0, j_1), j_1, \ldots, j_{k-1}, ch(j_{n-1}, j_n), j_n,$$

where

$$
\begin{aligned}
j_0 &\equiv P, \\
j_n &\equiv Q, \\
j_{i-1} &= \alpha(j_i),\ 1 \leqslant i \leqslant n \\
ch(j_{i-1}, j_i) &= \text{the shortest path from terminal } j_{i-1} \text{ to terminal } j_i \\
&\quad \text{in } G_s,\ 1 \leqslant i \leqslant n.
\end{aligned}
$$

We examine first how the $c(i, j)$, $(i, j) \in A_l$, are computed.

Consider a service $a_h \in A_s$. Between its origin j_0 and destination j_n, it stops at certain terminals $j_1, j_2, \ldots, j_{n-1}$. Between two consecutive stops j_i, j_{i+1}, the service travels nonstop on a path $p^h_{j_i j_{i+1}}$ called a service leg.

Every such service leg corresponds in fact to an arc of the service network G_s. Therefore, the service sequence $ch(i, j)$ between two terminals $i, j \in N$, may be found by a shortest path algorithm on G_s where the unitary cost of an arc p^h_{ij} of G_s is given by the following:

(i) the operation and delay costs due to over-the-road movements;
(ii) a penalty cost for excess assignment over capacity (if any);
(iii) a connection delay equal to (i) the delay due to stopping at a yard when the preceding arc belongs to the same service, or to (ii) the frequency delay when the preceding arc belongs to a different service.

These costs are all positive, so $c(i, j) \geqslant 0$ for all $(i, j) \in A_i$. As all terminal costs are also non-negative, negative cycling cannot occur in G_i.

On the other hand, the service network is connected by construction. So, $\tau(Q) = \infty$ only when the congestion on certain routes is so high as to prohibit the use of these routes for all or some types of services. But, this particular case may only occur during the initialization process and it is easily dealt with at that stage of the algorithm by, for example, uniformly decreasing the initial values of all the services using those routes. Consequently, the column generation procedure is well defined.

6. EXPERIMENTAL RESULTS

A version of the model has been developed to respond to the specific requirements of tactical planning in freight transportation by rail (for a detailed presentation of the fomulation, see Crainic *et al.*, 1984). Usually, several freight and freight train categories exist, each with very definite characteristics, so this is a complex problem where several modes compete for the same infrastructure and the full array of decision-making problems is present: service network design, multicommodity traffic routing, car classification at yards, train formation and length.

With this formulation, the algorithm presented in the preceding pages, as well as some strategies aimed at improving its efficiency, have been tested on data supplied by the Canadian National Railways (C.N. Rail).

The first data set contained traffic demand forecasts for the year 1990 for the western part of C.N.'s rail network. The traffic data is aggregated into 2613 final traffic classes of which the 735 largest (with a demand of at least 10 cars/week) count for over 95% of the total traffic demand. The aggregated physical network is made up of 107 nodes, of which 71 represent classification yards, the others being junctions and pick-up/delivery stations and 95 links.

Two different service networks were used in conjunction with this data. The first, called the "CANAT network", was supplied by C.N. Rail and was obtained through the company's CANAT system for the same data. (CANAT is a detailed, interactive, deterministic simulation system for the study of yard loads, train workloads and routings. It does not explicitly consider costs, delays, congestion or any other objective function whatsoever. The planner has to specify each train and its workload, based on his judgment and the status reports generated by the system. In this way, a solution is built and its effects are simulated.) The network has 415 service links generated by 14 rapid-high priority services and 177 normal speed-low priority services for a total of 191 services. Some preliminary results using this data and service network have also been presented in Crainic *et al.* (1984).

The second service network used in conjunction with the 1990 data set was generated by us (mainly for testing purposes), based on an empirical study of the O/D traffic demand matrix. This is a 399-services network, with 681 service-legs composed of 211 rapid services and 188 normal services. Passenger train services were also considered for all the experiments. The routes and frequencies were taken to be those appearing in the present schedules.

Tests were also conducted over a second set of data representing traffic demand forecasts for the year 1991 for the B.C. North Line, a subregion of the previous network. The physical network was much smaller: 41 links and 77 nodes, including 18 classification yards, 6 delivery/pick-up stations and 53 junctions.

The traffic data were aggregated into 488 traffic classes of which the 281 largest (with a demand of at least 10 cars/week) represented 99.32% of the total traffic demand.

An extensive study was conducted on this data at C.N. Rail, using the CANAT system, and they supplied us with the resulting service network: 47 services (4 "unit" services, 3 rapid services, 40 normal services) generating 86 service legs. As part of a comparative study on the same data, we also generated a service network composed of 44 services with normal speed/priority (72 service legs).

All programs are in FORTRAN. The tests were carried out on the Université de Montréal's CDC Cyber 835-855 system. However, the C.P.U. times shown here are Cyber 173 (a slower machine, previously available at our University) equivalent times.

Several types of experiments have been conducted on these data sets.

(i) Evaluation of the "CANAT solution": generation of the optimal traffic distribution on the CANAT service network with CANAT's service level;

(ii) Full optimization: on either service network from an arbitrarily chosen (usually high) initial level of service;

(iii) Reoptimization: beginning with the solution obtained from one of the first two experiments, the service level is marginally modified (usually, a uniform increase, by a small amount, of all frequencies) and the new problem is solved with the previous traffic distribution as an initial solution.

We first examine the performance of the individual procedures.

The frequency modification procedure essentially computes $\Pi(\Delta)$, the vector of finite differences for all services for a given step size Δ, finds the highest and the second disjoint-highest differences and modifies the chosen frequencies. Table 1 displays average solution times for this procedure on the various service networks.

As mentioned already, the performance of the procedure is determined by the size and intricacy of the service network and by the form of the physical network. These factors also influence the number of frequencies that can be efficiently modified during the execution of the procedure. In this sense, the results of Table 1 are relevant to networks with characteristics similar to those of the C.N. Rail network, i.e. quasilinear with few sections of the network permitting several parallel routes. The procedure might be made more efficient, if one is ready

Table 1. Average C.P.U. times for the frequency modification procedure

Size of the service network		Average C.P.U. time (sec.)
# services	# service-legs	
47	86	.11
191	415	.66
399	681	.94

Table 2. Impact of the number of repetitive executions of the frequency modification procedure on the total solution time and on the objective function value

Number of repetitions	Total solution time (sec. C.P.U.) (difference in %)	Difference in the value of the objective function (in %)
10	974.2 (+13.88)	+0.054
20	855.5 (-)	-----
30	791,0 (-7.54%)	-0.011

to pay the price of a large additional memory space, by the utilization of an incidence matrix indicating disjoint services. In this case, the procedure would be equally efficient for more general types of networks and would be able to decrease more than two frequencies at each occurrence. In that sense, the results of Table 1 also provide a good indication of the general behavior of the procedure.

To accelerate the solution procedure, we used a strategy which executes the frequency modification procedure several times between any two consecutive reoptimizations of the traffic distribution. This strategy proved to be very effective, especially in the last stages of the algorithm when a unitary step size Δ is used (note that, in our case, these "last" stages account for about half the total solution time). Yet, this has no discernible adverse effects on the quality of the final solution. These assertions are substantiated by the figures appearing in Table 2.

These results were obtained by solving the same problem, from the same initial solution with the same algorithm, while varying the number of repetitive executions of the frequency modification procedure between two consecutive traffic distribution reoptimizations. Twenty repetitions were then used as a standard in our programs.

Reoptimizing the traffic distribution is the most time-consuming part of the algorithm. This is due to the large number of subproblems that have to be solved at each iteration and to the fact that, for each subproblem, the best "new" itinerary has to be found prior to the traffic distribution reoptimization. On an individual basis however, these procedures are quite efficient as reflected by the figures in Table 3.

In Table 3, the average C.P.U. time for the itinerary generation procedure does not include the time necessary to find the shortest service paths (these times appear in Table 4), but does include the validation time (i.e., checking that this is a new itinerary). The time necessary to solve one subproblem thus depends on the size of the service network and on the average number of variables for each problem. Note, however, that this last quantity, as it appears in the second column of Table 3, represents the average number of itineraries generated for a subproblem during the entire solution process and not the number present at each subproblem reoptimization.

As described, the itinerary generation algorithm implies the generation of the shortest service paths each time a new variable has to be found. This is, however, much too inefficient. It requires very long solution times while the shortest service paths do not generally, nor greatly, change after the reoptimization of each subproblem. (It may be noted that, in the present formulation, the major impact of a reorganization of the car flow on the costs used to find the shortest service paths is via the capacity penalty terms, which may be activated or deactivated through traffic exchanges). The compromise reached was as follows.

Table 3. Average C.P.U. times to solve one reoptimization subproblem

Size of the service network (services/s-legs)	Average number of itineraries generated per subproblem	Average C.P.U. time (sec.)		
		Itinerary generation	Traffic dist. reoptimization	Total
47/86	2	0.008	0.02	0.030
191/415	6	0.038	0.04	0.085
399/681	12	0.033	0.08	0.120
399/681	15	0.056	0.11	0.172

Table 4. Average C.P.U. times to find the shortest paths in the service network for all pairs of yards

Size of the service network		Average C.P.U. times (sec.)
# services	# service legs	
47	86	0.20
191	415	2.14
399	681	2.80

(i) Split up the traffic classes into two groups, "large" and "small". Large traffic classes have over 90 cars/week for a maximum train length of 100 cars; there are 291 such traffic classes in the first data set, representing almost 80% of the total traffic demand;

(ii) Compute the new shortest service paths only after an execution of the service network design modification subproblem or after one solution cycle through either one of these two groups of traffic classes.

Once we had created these two groups and observed their relative weights, other accelerating heuristics became obvious. We used two, namely,

(i) solve less often for the small traffic classes;

(ii) use, for the small traffic classes, a "by origin" itinerary generation procedure. By origin implies that, for each yard, the tree of itineraries towards all yards is generated and these itineraries are used throughout that particular solution cycle for all small traffic classes originating at the yard; we used this strategy only a few times, at the beginning of the solution process.

These strategies helped to keep the solution time under control, without any real loss of accuracy. Table 5 summarizes the behavior and performance of the global algorithm through "typical" solutions of the following problems:

Table 5. Typical performance characteristics of the algorithm on various test problems

	Prob. 1	Prob. 2	Prob. 3	Prob. 4	Prob. 5
Shortest service paths					
# calls	14.0	27.0	46.0	45.0	25.0
# trees	994.0	1917.0	3266.0	3195.0	600.0
# paths	70574.0	136107.0	231886.0	226845.0	14400.0
Time	31.8	55.9	125.8	124.0	5.2
Frequency modification					
# calls	0	63.0	403.0	146.0	10.0
# computed Π	0	12033.0	105382.0	58254.0	430.0
Time	0	39.9	380.4	133.3	1.0
Traffic distribution					
Total time	391.3	596.0	1558.9	2276.7	104.8
Itinerary generation					
total gen.	2095.0	4540.0	8412.0	10793.0	708.0
new gen.	2095.0	4540.0	8412.0	2381.0	708.0
# trees	3356.0	4212.0	5928.0	10096.0	2462.0
# paths	22536.0	37392.0	46318.0	28856.0	7131.0
Time	244.7	266.0	295.0	518.9	21.1
Reoptimization					
# calls (cycles)					
Time	128.7	276.6	1024.0	1442.8	55.3
large t.-c.	4(7)	9(11)	11(19)	15(21)	9(12)
small t.-c.	4(4)	7(8)	9(17)	10(16)	8(8)
Total time	479.6	773.4	2139.4	2517.7	126.0

* Times in sec. C.P.U.

1. Evaluation of the CANAT solution for the first data set (735 traffic classes, 191 services).
2. Reoptimization of the CANAT solution.
3. Optimization with the first data set (735 traffic classes) and the second service network (399 services).
4. Reoptimization of the solution obtained in problem 3.
5. Optimization with the second data set (281 traffic classes) and the corresponding CANAT service network (47 services), followed by a reoptimization of the solution.

These results include how often each procedure was performed ("# calls") and the total time (sec. C.P.U.) necessary to perform the procedure this number of times. For the traffic distribution reoptimization procedure, the total number ("# cycles") of attempted reoptimizations is given for each group of subproblems. Usually, one cycle is sufficient to reoptimize the traffic distribution, except when major changes (such as eliminating an unproductive service which still carries a small amount of traffic) in the service network are introduced. For the itinerary generation procedure, "total gen." represents the total number of itineraries generated and stored at the end of the solution process. "New gen" is generally equal to this number, except for the fourth problem where it represents the number of new (with respect to those generated in problem 3) itineraries generated and stored during the solution process, while "total gen." sums up the generating effort of both the third and fourth problems.

The solution times shown in Table 5 are a good indication of the computing effort necessary to solve each type of problem. They are also extremely satisfactory considering the complexity of the problem solved and the effort required to solve the same problem by traditional methods.

As for the quality of the solutions obtained by the algorithm, the figures of Table 6 offer good indications. These figures represent the improvements (in percentages) obtained with respect to certain important system performance measures, relative to the respective CANAT solutions.

We do not intend to give here a lengthy analysis of the different solutions nor of the differences that may be observed between them. This is done elsewhere (Crainic, 1984). A few general characteristics of the method should however be stressed.

When starting from a "good" solution, the algorithm obtains significant reductions in system performance measures, especially those that are service related. When a less good initial solution is used instead, the algorithm gives less impressive results but, still no worse that those obtained by traditional methods.

In this respect, and noting that the formulation and the algorithm do not guarantee an optimum solution in the mathematical programming sense, it is advisable to marginally perturb and reoptimize any given solution. As illustrated in the fourth experiment, and verified by us in a number of other tests, significant improvements may be obtained by such a procedure. It is also a good investigative strategy for alternate solutions.

Due to the particular cost structure of the rail industry (emphasis on train costs and poor estimation of commodity-related delay costs), the major optimization effort has been directed toward the service network even at the expense of some losses in car performance. We think

Table 6. Improvements (in %) in certain system performance measures with respect to CANAT solutions

System performance measures	Prob. 2	Prob. 3	Prob. 4	Prob. 5
Total cost	5.48	-0.35	1.01	6.90
Tons-miles	0.37	-0.38	-0.39	2.65
Car-miles	0.33	-0.38	-0.38	1.75
Train-miles	12.82	13.82	15.58	11.89
Train costs				
delay	11.93	3.29	5.44	13.33
operation	11.55	3.65	5.76	12.48
total	11.63	3.57	5.69	12.67
Car costs				
delay	0.48	-15.46	-11.90	-1.73
operation	0.28	-1.38	-1.20	2.02
total	0.31	-3.64	-2.92	1.35

that a careful reevaluation of the cost structure, especially of commodity delay costs, might modify this pattern.

This underlines the relatively large amount of data necessary to calibrate and use the model: operating costs for each facility and each mode, delay costs for carriers and especially for commodities, operating characteristics for the physical and service networks, congestion (delay) functions for the nodes and links of the network, etc. In our opinion, however, the advantages of the method far outweigh these requirements.

7. CONCLUSION

We have presented a general modeling framework for the service network design problem for multimode multicommodity freight transportation in the case of a single authority controling both the service network and the movements of goods.

We have also described, in some detail, an efficient algorithm for solving this problem and we have presented results and performance measures drawn from experiments on a large rail application.

The modularity and versatility of both model and algorithm, as well as the quality of the results and the good performance of the algorithm, indicate that this is a method which may help to improve the planning process for freight transportation. The successful application (Roy, 1984) of this approach for the tactical planning of the "less-than-truckload" problem for a large, national, Canadian truck transportation company (which also uses the TOFC mode) is a further indication of the value of the method we propose.

We are presently continuing our work in the rail context and plan to apply this methodology to Brazilian freight transportation by rail, as well as for the study of some aspects of the air cargo transportation problem.

Acknowledgements—We want to thank Réjean Lessard for his excellent work of programming the algorithm. We would also like to express our appreciation to the anonymous reviewers for their valuable suggestions.

REFERENCES

Assad A. A. (1980a) Models for rail transportation. *Transpn Res.-A* **14A**, 205–220.

Assad A. A. (1980b) Models of rail networks: toward a routing/make-up model, *Transpn Res.-B* **14B**, 101–114.

Bazarra M. S. and Shetty C. M. (1979) *Nonlinear Programming*. John Wiley and Sons, New York.

Bodin L. D., Golden B. L., Schuster A. D. and Romig W. (1980) A model for the blocking of trains. *Transpn Res.-B* **14B**, 115–120.

Bronzini M. S. (1980) Evolution of a multimodal freight transportation network model. In Proceedings of the 21st Annual Meeting, Transportation Research Forum.

CACI Inc. (1980) Transportation flow analysis: the national energy transportation study (NETS), 3 vols. U.S. D.O.T. Report No. DOT-OST-P-10 (29–32).

Coutu J.-Y. (1978) Prédiction du trafic de wagons vides dans le réseau ferroviaire du C.N. à l'aide d'un modèle de gravité. Publication No. 118, Centre de recherche sur les transports, Université de Montréal.

Crainic T. G. (1982) Un modèle de planification tactique pour le transport ferroviaire des marchandises. Ph.D. thesis, Département d'Informatique et de Recherche Opérationnelle, Université de Montréal.

Crainic T. G., Ferland J. A. and Rousseau J.-M. (1984a) A tactical planning model for rail freight transportation. *Transpn Sci.* **18**, (2), 165–184.

Crainic T. G., Nguyen S. and Picard G. (1984b) Selected review of methodologies in freight transportation. Publication No. 333, Centre de Recherche sur les Transports, Université de Montréal.

Crainic T. G. (1984) A comparison of two methods for tactical planning in rail freight transportation. In *Operational Research '84* (Edited by J. P. Brans) Elsevier Publishers (North-Holland), Amsterdam, 707–720.

Fisk C. S. and Boyce D. E. (1983) Optimal transportation systems planning with integrated supply and demand models. Publication No. 16, Transportation Planning Group, Department of Civil Engineering, University of Illinois at Urbana–Champaign.

Friesz T. L., Gottfried J., Brooks R. E., Zielen A. J., Tobin R. and Meleski S. A. (1981) The northeast regional environmental impact study: theory, validation and application of a freight network equilibrium model. Report No. ANL/ES-120, Argonne National Laboratory, Argonne, IL.

Friesz T. L., Tobin R. L., Harker P. T. (1983) Predictive intercity freight network models: the state of the art. *Transpn Res.-A* **17A**(2), 409–417.

Friesz T. L., Viton P. A. and Tobin R. L. (1985) Economic and computational aspects of freight network equilibrium: a synthesis. *J. Regional Sci.* **25** (forthcoming).

Harker P. T. (1983) Prediction of intercity freight flows: theory and application of a generalized spatial price equilibrium model. Ph.D. thesis, Department of Civil Engineering, University of Pennsylvania, Philadelphia, Pennsylvania.

Kresge D. T. and Roberts P. O. (1971) Systems analysis and simulation models. In *Techniques of Transport Planning* (Edited by John R. Meyers) Vol. 2. Brookings Institute, Washington, D.C.

McGinnis L. F., Sharp G. P. and David Yu H. C. (1981) Procedures for multi-state, multi-mode analysis: Vol. IV, Transportation modeling and analysis. U.S. D.O.T. Report No. DOT-OST-80050-17/V.IV.

Peterson *et al.* (1975) The railcar network model (Edited by E. R. Petersen and H. V. Fullerton). Report No. 75-11, Queen's University, Kingston, Ontario.

Powell W. B. (1982) Analysis of airline operating strategies under stochastic demand. *Transpn Res.-B* **16B**, 31–43.

Powell W. B. and Sheffi Y. (1983) The load planning problem of LTL motor carriers: problem description and a proposed solution approach. Report No. EES-83-4, School of Engineering and Applied Science, Princeton University.

Roy J. (1984) Un modèle de planification globale pour le transport routier des marchandises. Ph.D. thesis, Université de Montréal.

Soumis F., Ferland J. D. and Rousseau J.-M. (1980) A model for large-scale aircraft routing and scheduling problems. *Transpn Res.-B* **14B**, 191–202.

Soumis F., Ferland J.-A. and Rousseau J.-M. (1981) MAPUM: a model for assigning passengers to a flight schedule. *Transpn Res. A* **15A**, 155–162.

Thomet M. A. (1971a) A combinatorial search approach to the freight scheduling problem. Ph.D. thesis, Department of Electrical Engineering, Carnegie–Mellon University.

Thomet M. A. (1971b) A user-oriented freight railroad operating policy. *IEEE Transactions on Systems, Man and Cybernetics*, Report No. SMC-1, 4, 349–356.

Vecellio R. L., Jones, P. S. and Day M.-D. (1981) Procedures for multistate, multi-mode analysis—Vol. VI: Transportation technology. U.S. D.O.T. Report No. DOT-OST-80050-17/V.VI.