

**A STRATEGIC MODEL FOR FREIGHT OPERATIONS ON
RAIL TRANSPORTATION SYSTEMS**

J. Enrique Fernández L., Joaquín de Cea Ch.
Depto. Ingeniería de Transporte, U. Católica de Chile
Casilla 306, Santiago 22, CHILE
FAX: (56-2) 686 4818; e-mail: jef@ing.puc.cl

&
Ricardo Giesen E.
Fernández y De Cea Ingenieros
Lota 2257, Of. 402, Santiago, Chile
FAX: (56-2) 234 1578; e-mail: rgiesen@FDCconsult.com

ABSTRACT

In this work we present a new strategic model for freight rail transportation systems. The objective of the model is predicting equilibrium flows and level of services, given O-D transportation demands for a set of different products. Programmed rail services, corresponding to trains that run over the rail network with fix schedules are considered. The model proposed has the following distinctive characteristics: i) Railway services and their operational characteristics are explicitly considered. ii) The distribution of empty cars and its assignment over the rail network is treated simultaneously with the assignment of products to be transported, considering the operating costs over the network and the transportation priorities. iii) Different priorities are assigned to different types of cars to be moved depending on the products transported. iv) Capacity constraints are considered for the movement of different products, depending on the availability of the type of cars necessary and the demands of products competing for the use of the same cars. This constraint makes that some shipments must wait longer in the origin or transfer yard if not enough capacity exists. A mathematical formulation of the problem described is developed and its characteristics analyzed. Finally, possible solution approaches are discussed.

1. INTRODUCTION

Given the importance and distinctive characteristics of railroads, it is necessary to develop good representations of them for the strategic analysis of interurban freight transportation systems. The objective of this work is the development of a strategic model for a freight railway system operating in a given region. Its output must be the level of service provided by the railway system considered: travel and waiting times, flows of shipments by type of product and rail cars movements for each type of car considered. Such result must be obtained using as inputs: the transportation O-D demands for the products considered, the characteristics of the railway system and its components (yards, lines, etc.), the operational policies and services offered and the products priorities. The model must be able to be incorporated to a strategic multi-modal freight network model (Fernández et al. 1998).

The model takes a short run approach and assumes that all the rail services offered are known and given. The network proposed includes several small sub-networks to represent yards operations and the concept of route sections is used to represent line operations (De Cea and Fernández, 1993).

The paper is divided in six parts: After this introduction, in section 2 an analysis of freight rail operations modeling is made, and the corresponding literature is reviewed; in section 3, the model structure and car demands are defined and the main variables and behavioral assumptions are specified; section 4 defines de network representation and flow delay functions; in section 5 the mathematical representation of the problem is given and its characteristics analyzed; in section 6 solutions approaches are discussed; finally, section 7 contains the main conclusions.

2. MODELLING FREIGHT OPERATIONS ON RAIL SYSTEMS

Rail operations are performed over a railroad network that includes classification yards for trains formation, many small stations for secondary operations of trains and cars, a set of primary and secondary rail lines and a set of junction points. Rail cars move over the lines in groups called blocks, that are considered as transport units from an operational point of view; one or more blocks plus an engine form a train. Trains travel over predefined routes with specific frequencies and itineraries.

Usually, over the same rail lines different types of services operate: freight and passenger trains which can also be divided in different types of services; therefore, special priority rules must be used to avoid operating conflicts. In order to guaranty a safe operation the line is divided in several line segments, each of which can be used only by one train at the same time. The operation is guided by a signaling system that allows imposing constraints to trains operations, such as speed limits for different line segments or stop and temporal resting requirements at some points. The length of line segments depends on the characteristics of the signaling system.

While operating over a line, a train can overtake or being overtaken by other trains operating at different speeds; in the case of single track operations it can also cross with trains operating in the opposite direction. In all these cases trains with lower priority will yield to those with higher priority; they will have to stay on a siding line until the train with higher priority has emptied the corresponding line segment.

Rail cars, which constitute the basic operating unit in a rail system, are subject to different operations within their operating cycle: i) the client or shipper asks for a car to load his shipment; ii) the carrier (rail operator) selects a car with the required characteristics from the nearest yard possible, which after being inspected is sent to the shipper that loads it

(usually inside his own precinct); iii) the loaded car is sent to the origin yard, where it is classified and grouped in a block; iv) the block is assigned to a train with a given itinerary; v) when the train stops in a yard the cars and engines are usually inspected, also blocks can be transferred to other trains and if the yard is the destination for the block, the cars are separated and classified again; vi) cars which have arrived to their final destination are sent to a station for unloading; vii) finally, empty cars are inspected and cleaned after what they are available for a new assignment.

Given the characteristics of rail operations, cars spend most of their time inside classification yards (77% in US systems, Reebie Associates, 1972). Therefore, the understanding and adequate representation of yard operations is of fundamental importance for the correct analysis of rail freight operations.

A train that enters a station usually undergoes a sequence of processes, that can be classified in four general tasks: i) train reception and inbound inspection; ii) cars classification and block formation; iii) connection delay and train assembly; iv) outbound inspection and departure. According to works developed by Folk (1972) and Reid et al (1972), for the U.S. freight railways, the classification step is the main source of delays for a car going through a station. They also found that the delays related to operation (iii) are mainly a consequence of the trains exit programming (or train scheduling), rather than of the limited capacities for train assembling. Petersen (1977a) concluded that operations (i) and (iv) do not generally constitute bottlenecks in yard operations and therefore can be adequately represented by fix times.

2.1 Yards Types and Yard Operations Modeling

Railway systems use different types of yards. Petersen (1977a) classifies them in five categories: i) Simple yards; ii) single ended flat yards; iii) doubled ended flat yards; iv) directional flat yards; v) hump yards. The first four categories are flat configurations of lines and the main differences among them are related to the existence of different groups of lines assigned to arrivals, departures and classification of cars and the use of special equipment (yard engines) for the operations. The last category corresponds to the larger and most complex type of yard; there, the classification and assembly operations are always separated and therefore must be represented independently.

Yard production functions are aggregated relations between flows or yard throughput, yard configuration (capacity) and delays; they are mainly used with network models and can be oriented to operational, tactical (Thomet, 1971; Petersen and Fullerton, 1975; Assad, 1980; Crainic et al, 1984; Marin et al, 1993), or strategic analysis (Crainic, Florian and Léal, 1990). In most present models, a yard is commonly represented by only one network arc (Crainic et al, 1990), with the initial node receiving the traffic and the final node dispatching trains.

2.2 Modeling Line Delays

Following Petersen and Taylor (1982), delay models developed can be classified in three classes: simulation, optimization and analytical models. Analytical models are in general the most appropriate for being used in strategic models. They are based on analytical expressions that can be used to obtain the value of delays, for different types of trains, as a function of the number and type of trains operated, the operational policies and the line configuration. Some models proposed in the literature (Petersen, 1974, 1975; English, 1977) that take into account several line operational variables, give good results when the level of congestion is low or intermediate, but tend to overestimate delays when congestion is significant. In such cases, cars accumulate inside yards waiting for trains being

dispatched when lines capacity allows it; therefore, excess delays due to lines congestion are experienced inside yards. To avoid these problems, Franck, (1966) and Petersen (1977c) develop expressions that incorporate connection times inside yards.

In network models three main approaches have been used: i) constant delays (Payraud, 1981), in which case congestion is not considered; ii) a “black box ” sub model, to calculate delays (Petersen y Fullerton, 1975); the black box is some external delay model that is called each time delays must be calculated; iii) delay functions with approximated parameters (Crainic, 1984): analytical approximated expressions, calibrated on the basis of historical values, are used.

2.3 Empty Vehicles Modeling

Consideration of empty cars is of fundamental importance in modeling freight railway systems. Mendiratta, (1982) found that in the U.S. railway system cars spend 40% of their time traveling empty. Moving empty vehicles is necessary in order to compensate non-balanced traffics and it is influenced by cars specialization in specific loads or products, operational constraints and regulations (Dejax and Crainic, 1987).

It is in general necessary estimating the demand for empty cars required at different points of the rail network, to move the shipments transported by the railway system. It is also necessary knowing the number of empty cars available at different points at a given time. Different methods to consider empty vehicles movements, in freight transportation network models, have been proposed in the literature (Dejax and Crainic, 1987). Coutou (1978) and Crainic et al. (1990), used gravitational models to estimate the flows of empty rail cars.

STAN (Crainic et al., 1990) is the only strategic freight transportation model that explicitly considers empty vehicles for the railway system. Empty railcars are considered as a different product, that must be transported to satisfy specific demands between O-D pairs; however only one type of empty car is considered, which is defined as a combination of the cars used in the rail system, with weights proportional to the number of each of them required to transport each of the products.

2.4 Strategic Models and Railway Systems Representation

The most important strategic models available in the literature, for interurban transportation systems, are FNEM (Freight Network Equilibrium Model, Friesz et al, 1986) and STAN (Strategic Transportation Analysis, Crainic, et al, 1990). They represent significant advances for interurban freight network transportation systems modeling. However in FNEM there is no explicit consideration of the special characteristics of railway systems; therefore modeling of railway operations must be made through the specification of appropriate cost functions for the network elements. Therefore, there is no possibility of considering special features or variables corresponding to rail systems.

STAN, uses a representation of railway operations, taking into account several distinctive characteristics of this mode, like the (simplified) consideration of empty cars and the explicit (aggregate) representation of yards. Yard representation is simplified by using only one arc where polynomial flow delay functions, depending on the total number of cars demanding the use of the yard are used; therefore, only the total average delay for the yard is considered. There is no explicit representation of different services (different types of trains, frequencies and itineraries), which makes impossible a representation of connection delays. Arcs that connect different yards represent line operations; they are associated with flow delay functions that represent flows as a weighted sum of the cars moving through the arc.

3. MODEL STRUCTURE, MAIN VARIABLES AND ASSUMPTIONS

3.1 Basic Definitions

Given that we will be frequently using the terms service, itinerary, route, service section and route section, we now define what we mean by each of them.

A “train service” or just a “service”, is a group of engines that run back and forth between two yards on the rail network; each of them corresponds to what we call a train. All the trains in the same service have identical capacities and operating characteristics and travel on the network going always through the same sequence of lines (network links) and yards (nodes), referred to as an “itinerary”.

We assume that each engine stops at each yard included in its itinerary, to allow blocks and/or cars to be connected or disconnected from the train; therefore, those yards where services do not stop, do not need to be included in the description of an itinerary. Each train service is defined by its itinerary, and the train frequency and capacity.

A “service section” is any portion of a service between two not necessarily consecutive yards of its itinerary, and a “service segment” represents a portion of a service between two consecutive yards in its itinerary (notice that these yards are not necessarily physically consecutive in the real network).

A “railcar route” is any path that a car can follow on the service network in order to travel between any pair of yards. In general it will be identified by a sequence of yards, the first one being the origin of the trip, the final being the destination and all the intermediate yards representing transfer points (from one train to another). The portion of a route between two consecutive transfer yards is called “route section”.

Each route section contains all service sections that a given car or block can use to travel between a pair of yards. In some cases, some service that in principle could be used has associated a too large travel time and therefore is not attractive to be used. Therefore, in order to define which is the set of “attractive” services, to be used between a given pair of yards, it is required that each of the services included contributes to reduce the expected travel time (De Cea and Fernández, 1993). In the particular case that all services available have the same travel time, the attractive set will contain all of them. However if travel times are different it is necessary to find the sub-set of services that minimize the expected travel time.

3.2 Model Variables and Behavioral Assumptions

In the following we will assume a network represented by a graph $G(N, A)$, with node set N and link set A . The set N is composed by two types of nodes: i) Centroids, which are points where shipments origins and destinations are supposed to be concentrated; ii) other nodes, representing yard nodes, stations and junctions. On the other hand, the link set, A , is composed by three types of links: i) Consolidation links, Co , which connect a centroid to the yards that can serve it; ii) Classification links, Cl , which represent yard classification operations; iii) Route sections links, S , which represent connection and in line delays (a more detailed explanation is given in points 4.2 and 4.3).

For the network characteristics we will denote:

W : Set of network origin-destination (O-D) pairs.

w : An element of set W .

\overline{W} : Set of network pairs of yards.

\overline{w} : An element of set \overline{W} . Each O-D pair, $w \in W$, has associated only one pair of yards, $\overline{w} \in \overline{W}$.

- R : The set of route in $G(N, A)$ available to cars trips.
- R_w : Set of feasible routes associated with O-D pair w .
- r : An element of set R .
- f_t : Frequency of train service t . It is measured in numbers of trains during the analysis period.
- k_t^s : Capacity of train service t on service section s , measured in Tons.
- f_s : Total frequency on route section s ($frec_s = \sum_{t \in s} frec_t$). It is measured in number of trains during the analysis period.
- K_s : Total capacity on route section s ($K_s = \sum_{t \in s} k_t^s$), measured in Tons.
- $h_r^{vp,\pi}$: Flow on route r of cars type v loaded with product π (that has associated a priority p), measured in cars during the analysis period.
- $h_r^{vp,\varepsilon}$: Flow on route r of empty cars type v (that has associated a priority p), measured in cars during the analysis period.
- U_w^{vp,π^*} : O-D level of service (composite cost) experienced over all used routes between pair w , by cars type v , carrying product π , with priority p .

We assume that the rail freight system offers a set of known services, corresponding to trains that travel with programmed schedules, following specific routes over the rail network; however, the number and type of cars that these trains carry will depend on the specific transport demands that the system faces. The operation of freight rail systems with fix service schedules, is today becoming more generalized in order to be able to offer a good and reliable level of service to shippers (Jovanovic and Harker, 1991; Bouley, 1987). The main variables of the model will be the flows for each type of car considered and the travel times experienced, including train movements on rail lines and cars operations within yards. Train travel times between yards will be considered fix, given that line operations are programmed; therefore, waiting times on sidings due to overtaking operations by trains with higher priority are programmed and included in the trains travel times between yards.

Once a given shipment has been loaded on an appropriate car and placed in the origin yard, it can be assigned to several different services going out of the yard. These can either go to an intermediate yard, where it will be transferred to other service (train), or directly to its final destination (if such direct service exists). The routing decisions are taken by the rail operator with the objective of minimizing travel times, given the constraints of the operation, the priorities assigned to each product and taking into account the level of service perceived by the shippers. Each service has a capacity over each service segment, which depends on the engine power and the line operating characteristics, where the main constrains corresponds to the slope of the line. Therefore on each service segment a given service can carry a maximum number of Tons (including cars plus load weight) (Hay, 1982).

In order to define the capacity available in a given service at a given yard we make the following assumptions:

- i) Cars carried by a train that is coming into a yard have priority to continue in the same train over cars waiting in the yard. Therefore, the capacity of a service available at a given yard corresponds to the total capacity of the corresponding train, less the number of cars already in the train coming from a preceding yard and continuing to a next yard on the service itinerary. This assumption is similar to that used by Crainic and Gendreau (1986) to analyze connection delays.

ii) Cars waiting in a yard, to be connected to a train, will be processed taking into account the priorities assigned to the products carried.

The costs considered by rail operators in the route choice process over their rail network, are represented by the following composite cost function over route r , for cars type v , loaded with product π , that has priority p .

$$C_r^{vp,\pi} = OC_r^{vp,\pi} + \phi^v \cdot tmg_r^{vp} \quad (1)$$

Where, $OC_r^{vp,\pi}$ represents the operating cost associated to transport a car type v , loaded with product π , over the route r (which is equal to the sum of operating costs, $OC_a^{vp,\pi}$, experienced on all the arcs used by route r); tmg_r^{vp} corresponds to the marginal travel time experienced by the rail operator and ϕ^v is the corresponding time value. We assume that arc operating costs are constant, independent of flows; they depend on the arc operating characteristics (gauge, slope, curvatures, state of lines etc.) for line operations and on the operations performed at yards. Therefore marginal operating cost is equal to average operating cost and both are represented by $OC_r^{vp,\pi}$ in expression (1). However, travel times depend on yards congestion level and therefore tmg_r^{vp} will be an increasing function of flow.

The assignment of railcars flows over the railway network defined in section 4.1 will be determined by the following conditions:

$$C_r^{vp,\pi^*} \begin{cases} = U_w^{vp,\pi^*} & \text{if } h_r^{vp,\pi^*} > 0 \\ \geq U_w^{vp,\pi^*} & \text{if } h_r^{vp,\pi^*} = 0 \end{cases} ; \forall r \in R_w, \forall w \in W, \forall v, \forall p, \forall \pi \quad (2)$$

With h_r^{vp,π^*} and U_w^{vp,π^*} representing path flows and O-D level of services as defined before. Then, in a solution satisfying (2) all used rail routes (with positive flows) will experience the same marginal composite cost value: $C_r^{vp,\pi^*} = U_w^{vp,\pi^*}$; all routes without flow will present a higher cost. Notice that given the expression (1) this corresponds to a system optimum solution for the rail operator (second Wardrop principle, Wardrop, 1952).

3.4 Modeling of Cars Demands

Transportation demands expressed in Tons for each type of product π , among different pairs of yards w , Ton_w^π , are exogenously specified and given to the model; also the types of cars, v , available to transport each product, and product priority, p , are exogenously specified. Using these data, demands $T_w^{vp,\pi}$, for cars type v , with priority p , loaded with product π , between O-D pair w , are calculated using the following expression:

$$T_w^{vp,\pi} = \delta_{\pi-vp} \frac{Ton_w^\pi}{(\overline{\omega}_v^\pi - \overline{\omega}_v^\varepsilon)} \quad (3)$$

with:

$$\delta_{\pi-vp} = \begin{cases} 1 & \text{If the product } \pi \text{ is transported} \\ & \text{in cars type } v \text{ and has associated} \\ & \text{a priority } p. \\ 0 & \text{Otherwise,} \end{cases} ; \forall v \in V, \forall p, \forall \pi \quad (4)$$

where, $\bar{\omega}_v^\varepsilon$ represents the weight of empty cars type v , and $\bar{\omega}_v^\pi$ the weights of loaded cars v transporting product π . Notice that each product π is associated with only one car type, however a given car type can be assigned to carry more than one product. Equation (3) assumes that all loaded cars are full loaded with only one type of product.

Then, the number of empty cars, type v , supplied, $O_i^{vp,\varepsilon}$, and demanded, $D_i^{vp,\varepsilon}$, at each yard i can be calculated as the difference between, the number cars loaded with products which final destination is yard i and the number of cars needed to carry the products generated in the same yard.

$$O_i^{vp,\varepsilon} = \delta_{v-p} \cdot \text{Max} \left\{ 0, \sum_{k \in O(i)} \sum_{\pi} T_{(k,i)}^{vp,\pi} - \sum_{j \in D(i)} \sum_{\pi} T_{(i,j)}^{vp,\pi} \right\}; \forall i \in O, \forall v \in V, \forall p \quad (5)$$

$$D_j^{vp,\varepsilon} = \delta_{v-p} \cdot \text{Max} \left\{ 0, \sum_{k \in D(j)} \sum_{\pi} T_{(j,k)}^{vp,\pi} - \sum_{i \in O(j)} \sum_{\pi} T_{(i,j)}^{vp,\pi} \right\}; \forall j \in D, \forall v \in V, \forall p \quad (6)$$

$$\delta_{v-p} = \begin{cases} 1 & \text{If empty cars type } v \\ & \text{have priority } p. \\ 0 & \text{Otherwise.} \end{cases}; \forall v \in V, \forall p \quad (7)$$

Where $O(i)$ is the set of all origins for cars going to yard i , and $D(i)$ is the set of all destination yards for cars coming out of yard i . It is important to notice that (5) and (6) must be calculated before formulating the equilibrium problem, because the corresponding values of cars supplied and demanded enter to it as exogenous data (see equations (24) and (25) in section 5).

Then, empty cars O-D demands between pairs of yards, $T_w^\varepsilon = \{T_w^{vp,\varepsilon}\}$, are calculated using an entropic double constrained distribution model as follows:

$$T_w^{vp,\varepsilon} = A_i^{vp,\varepsilon} O_i^{vp,\varepsilon} B_j^{vp,\varepsilon} D_j^{vp,\varepsilon} \exp\left(-\left(\gamma \cdot C_w^{vp,\varepsilon*}\right)\right) \quad (8)$$

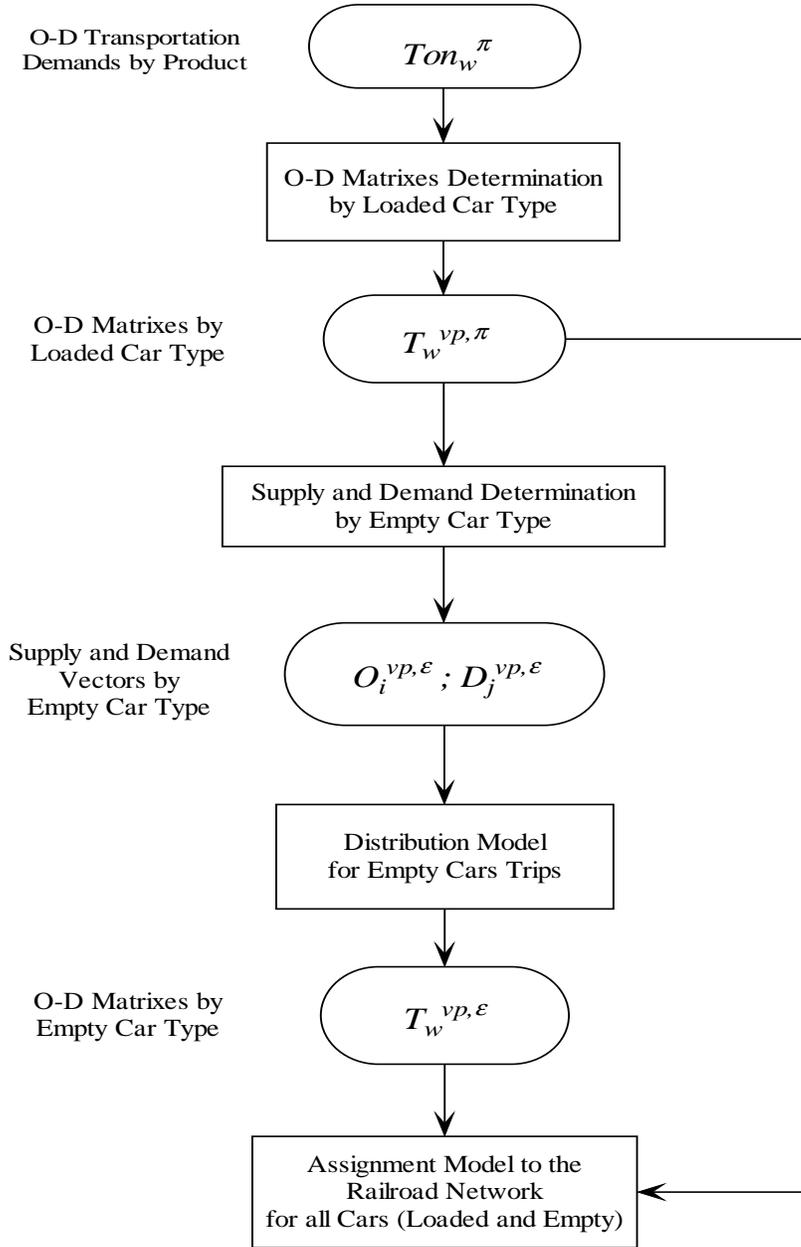
where the impedance, $C_w^{vp,\varepsilon}$, corresponds to the transportation cost between O-D pair w , for empty car type v , with priority p ; $A_i^{vp,\varepsilon}$ and $B_j^{vp,\varepsilon}$ are the usual balancing factors:

$$A_i^{vp,\varepsilon} = \frac{1}{\sum_j B_j^{vp,\varepsilon} D_j^{vp,\varepsilon} \exp\left\{-\left(\gamma \cdot C_w^{vp,\varepsilon*}\right)\right\}} \quad (9)$$

$$B_j^{vp,\varepsilon} = \frac{1}{\sum_i A_i^{vp,\varepsilon} O_i^{vp,\varepsilon} \exp\left\{-\left(\gamma \cdot C_w^{vp,\varepsilon*}\right)\right\}} \quad (10)$$

The different stages of the model, main variables and interrelations considered are shown in Figure 1.

Figure 1. Model Structure and Components



4. NETWORK REPRESENTATION AND FLOW DELAY FUNCTIONS

4.1 Line Operations

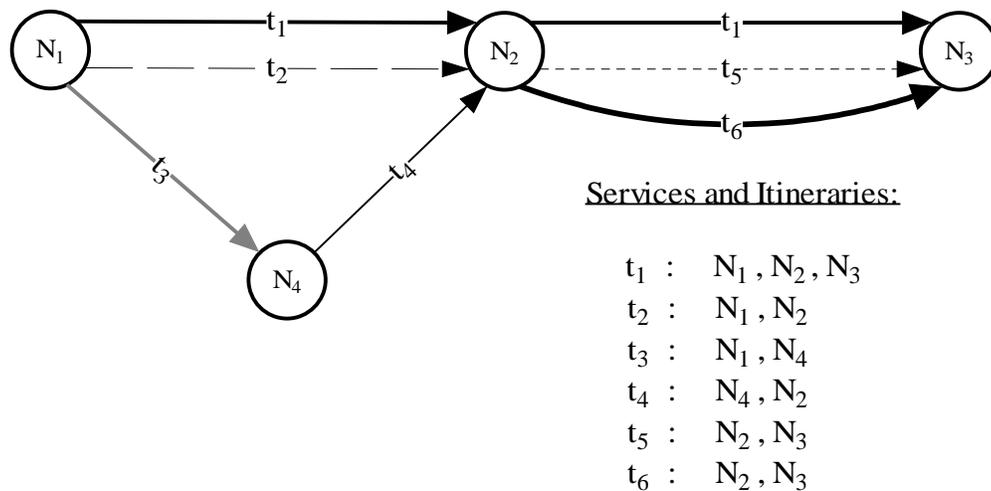
According with the assumptions defined above, the network must represent the different services offered, corresponding to all freight trains scheduled by the rail operator. Each of these services, $t \in T$, is defined by its origin yard, $O(t) \in N$; its destination yard, $D(t) \in N$; a itinerary from $O(t)$ to $D(t)$, defined by a sequence of service segments, each of them connecting two consecutive yards in the train itinerary; the travel time and carrying capacity on each of the service segments; a set of intermediate stops (yards), corresponding to transfer yards where the train can leave or take cars; and finally a service frequency.

Depending on the service characteristics, a train can cross some yards without stopping, therefore those yards do not need to be included in the train itinerary.

Figure 2 shows a graphical representation of a network of services, $G_t = (N, A_t)$, where services are specified over a representation of the infrastructure network (lines and yards). Nodes represent yards (considering in this case only one node by yard, to simplify the representation) and arcs represent service segments; parallel arcs between two common nodes represent services running over the same segment of rail line.

The network represented in Figure 2 corresponds to a rail system with four yards ($N_1, N_2,$

Figure 2. Graphical Representation of a Network of Transportation Services



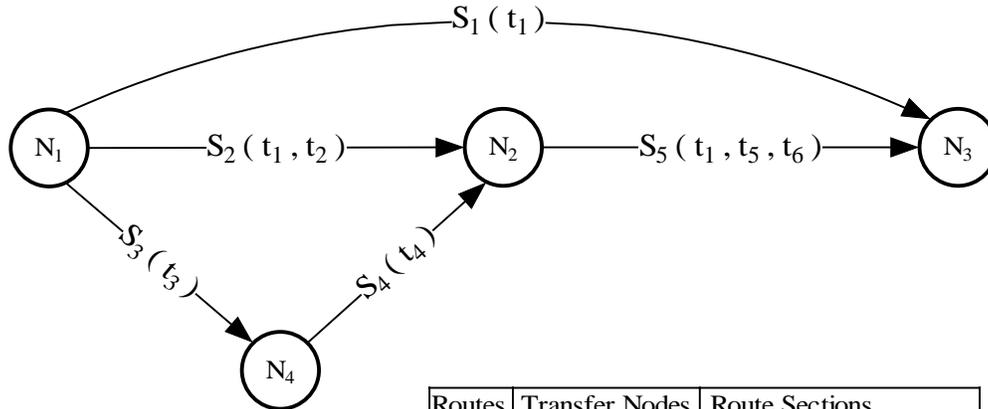
N_3 and N_4) and six different services (t_1, t_2, t_3, t_4, t_5 and t_6). The itineraries for each of them are indicated by the sequence of yards at which the corresponding train stops. From this network we can build a new auxiliary network as shown in Figure 3. This new network has virtual arcs corresponding to route sections; a route section is built choosing all “attractive” service sections that directly join a pair of yards. The set of “attractive” or “common” services between two yards corresponds to that subset that minimizes the total connection plus travel time, for a given car or block of cars. When all services running between a pair of yards have the same travel time (same train speed) as is in general the case for freight services, the attractive set will contain all the services available. However if services with different travel times exist, then the definition of the attractive set needs the solution of an hyperbolic problem (see, De Cea and Fernández, 1993).

For the example represented in Figure 3, five route sections are defined. For each route section, S_i , the corresponding set of attractive (or common) services is indicated in brackets. Notice that for route sections S_1, S_3 and S_4 , the attractive set contains only one service, because in those cases cars have only one direct service (without intermediate transfers between different services) available, to travel between the corresponding pair of yards. This network representation allows a consistent treatment of cars delays inside yards and the consideration of alternative services to carry cars from a given yard, as we will see later.

4.2 Yard Operations

As analyzed in Section 2, the most important operations to determine yard delays are car classification and car connection. Also, given that some type of cars can be in short supply, it is convenient considering delays from consolidation operations, corresponding to the

Figure 3. Auxiliary Network of Transportation Services



Routes	Transfer Nodes	Route Sections
R ₁	(N ₁ , N ₃)	S ₁ {(N ₁ , N ₃)→t ₁ }
R ₂	(N ₁ , N ₂ , N ₃)	S ₂ {(N ₁ , N ₂)→(t ₁ t ₂)}, S ₅ {(N ₂ , N ₃)→(t ₁ t ₅ t ₆)}
R ₃	(N ₁ , N ₄ , N ₂ , N ₃)	S ₃ {(N ₁ , N ₄)→t ₃ }, S ₄ {(N ₄ , N ₂)→t ₄ }, S ₅ {(N ₂ , N ₃)→(t ₁ t ₅ t ₆)}

process of transferring a load from the shipper storage area, to the car type required to transport it over the rail system. This process can experience delays, in addition to the time necessary to perform the car loading operation, if the needed type of car is not readily available when necessary. These delays are considered in consolidation arcs that also act as access arcs, connecting the centroids, to the rail network (usually to the entrance of some yard). Therefore railway yards are represented as small subnetworks as shown in Figure 4.

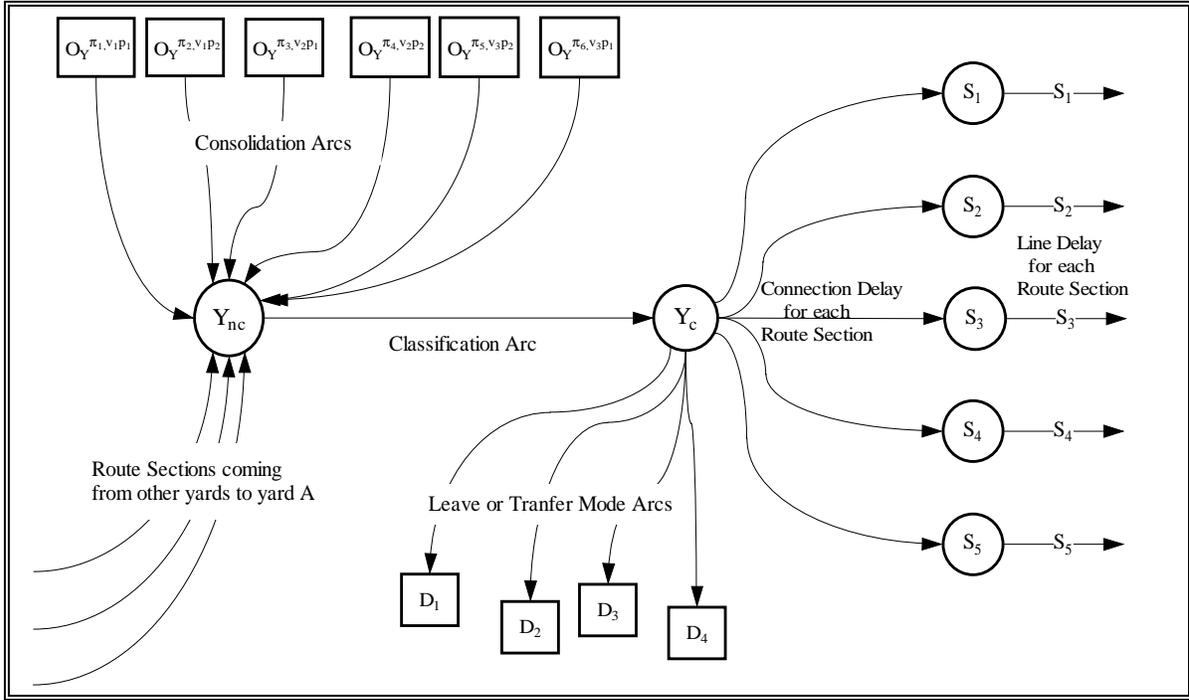
4.3 Arc Cost Functions

4.3.1 Consolidation Delays

A consolidation delay corresponds to the time elapsed between the moment at which a shipment is made available at the origin centroid and the moment at which it is completely loaded into the appropriate type of car, to be transported over the rail system. The total time spent can be divided in two parts: a fix part, corresponding to the time necessary to load the car, at the rate given by the technology used and a variable part, corresponding to the time that the shipment must wait until an appropriate type of car is made available for the loading operation.

The fix part is assumed known and given for each type of product at each yard; the variable part depends on the availability, at that moment, of the type of cars necessary. Therefore it will depend on the relation between the total supply of such cars in the system and the demands, during the analysis period, made by shipments with the same car requirements and higher or equal service priority.

Figure 4. Detail Network Yard Representation,



We use the following BPR specification to represent the average consolidation delay for a product π to be loaded on a car type v , which has associated a priority p , in yard i :

$$DMeCo_i^{vp, \pi} = DFCo_i^{v, \pi} + \beta Co_i^v \cdot \left(\frac{D^{vp}}{NC^v \cdot \tau} \right)^{nCo_i^v} \quad (11)$$

where: $D^{vp} = \sum_{q \geq p} \sum_{\bar{w} \in \bar{W}} T_{\bar{w}}^{vq} \cdot t_{\bar{w}}^{vq} = \sum_{q \geq p} \sum_{a \in \{CIYS\}} f_a^{vq} \cdot t_a^{vq}$ (12)

$$T_w^{vp} = T_w^{vp, \mathcal{E}} + \sum_{\pi} T_w^{vp, \pi} \quad (13)$$

with:

$DFCo_i^{v, \pi}$: Fix time necessary for loading product π , on a car type v , at yard i (hrs).

D^{vp} : Number of type v car-hrs, demanded in the system during the analysis period, by shipments with priority p or higher.

$T_{\bar{w}}^{vp}$: Number of cars type v , with priority p , demanded to travel between the pair of yards \bar{w} , during the analysis period. This variable is given by equation (13).

$t_{\bar{w}}^{vp}$: Expected travel time between the pair of yards, \bar{w} , for type v cars with priority p .

f_a^{vp} : Flow of cars type v , with priority p , running over arc a , during the analysis period. (for product carried on cars type v that has associated priority p)

t_a^{vp} : Travel time over arc a , for cars type v with priority p . It is expressed in hours.

NC^v : Fleet of type v cars available in the system.

τ : Length of the analysis period expressed in hours.

βCo_i^v & nCo_i^v : Calibration parameters.

This specification allows considering the constraint imposed by the existing number of cars of each type, available in the system. Expression (11) is based on the total availability of cars in the system during the analysis period, given that the operator can reassign cars to different yards in order to favor the movement of high priority shipments. Therefore, the availability of empty cars in a particular yard does not mean that they will be used to dispatch the shipments waiting in the yard, because those cars could be sent to other yard in order to serve shipments with higher priority.

Notice that according to expression (11) consolidation delays depend on the values of cars flows over all network, with exception of the same consolidation arcs. Therefore it is independent of its own flows, because consolidation arcs do not involve the use of cars. Then expression (11) will take a constant value, if diagonalized.

4.3.2 Classification Delays

To calculate classification delays we consider individual cars (Turnquist and Daskin, 1982; Crainic et al, 1984) and not trains as Petersen (1977b). Taking into account car delays, allows a more precise and reliable modeling of classification delays.

Instead of using a queuing formula (as Crainic et al, 1984), we also propose here the use of BPR type functions. Then, the average classification delay for a railcar in yard i will be given by the following expression:

$$DMeCl_i = DFCl_i + \beta Cl_i \left(\frac{f_i}{CAP_i} \right)^{nCl_i} \quad (14)$$

Notice that this delay is the same for all types of cars going through the yard.

$DFCl_i$: Average classification delay for a car in yard i , in free flow conditions (when there is no congestion). This is a technical parameter, which depends on the yard layout and classification equipment.

f_i : Flow of cars to be classified in yard i during the analysis period, expressed in cars. $f_i = \sum_{\pi} f_i^{vp,\pi}$.

CAP_i : Classification capacity of the yard i , expressed in cars, for the analysis period.

βCl_i & nCl_i : Calibration parameters.

4.3.3 Connection and Route Sections Travel Delays

According with the adopted yard network representation (Figure 4), there is one connection arc for each route section leaving a yard. Also, as was explained in section 3.2, travel times are assumed fix and known for each train running over each section of the rail line. Therefore, the section travel time and the connection delay can be considered together, adding both times and the separate representation of a connection arc can be eliminated.

The connection delay depends on the frequency of the common services included in the route section, the total capacity of the route section and the number of cars competing for it. Therefore, following a similar approach to that used for urban public transport services with capacity constraints (De Cea and Fernández, 1993), we will use the following function to calculate the average delay on a route section s , for cars with priority p :

$$DMeS_s^p(V_s^p) = DFS_s + \left(\frac{\alpha}{f_s} \right) + \beta S_s \cdot \left(\frac{V_s^p + \tilde{V}_s^p}{K_s} \right)^{nS_s} \quad (15)$$

$$\text{with : } \tilde{V}_s^p = \hat{V}_s^p + \check{V}_s^p \quad (16)$$

Where:

- DFS_s : Fix delay corresponding to the time taken by following operations: cars connection and outbound inspection in the origin yard; programmed travel time between yards; reception and inbound inspection in the destination yard.
- f_s : Total service frequency of route section s . Corresponds to the sum of frequencies over all common (or attractive) services in the route section.
- K_s : Capacity of route section s . Corresponds to the total Tons. of load that can be transported over route section s during the analysis period by all trains in the route section.
- α : Calibration parameter, associated to the probabilistic distribution of times between services.
- βS_s & nS_s : Calibration parameters related to the description of the connection congestion phenomena.
- V_s^p : Flow of cars with priority p , waiting to be connected to route section s services in yard $i(s)$. It is expressed in Tons. and corresponds to the total weight of the corresponding waiting cars.
- \tilde{V}_s^p : Complementary flow. It is expressed in Tons. and corresponds to the sum of competing plus preceding flows, as defined next.
- \hat{V}_s^p : Competing flow. Expressed in Tons. corresponds to those car flows, with the same priority p , that are waiting in the same yard, but are assigned to other route sections (different from s) that contain services also included in s .
- \check{V}_s^p : Preceding flow. Expressed in Tons., corresponds to the sum of car flows already coming in trains contained in route section s and that are not left in yard $i(s)$, plus car flows waiting for connection in the same yard $i(s)$, but with higher priority than p , and using route sections that include services also contained in s .

According with the definition of the variables included in (15), we are assuming (as Crainic and Gendrau,1986) that cars already connected to a train coming into a yard, have priority to continue in the same service (connected to the same train) beyond that yard. Therefore, cars waiting for a route section at a given yard, have only available the remaining capacity of the services coming into the yard. Such remaining capacity includes that produced by cars disconnected at that yard (either because final destination or transferring point). Then, we can calculate competing, preceding and complementary flows as follows:

$$\hat{V}_s^p = \sum_{t \in s} \sum_{r \in S_{i(s)}^+} (v_r^t)^p \quad (17)$$

$$\check{V}_s^p = \sum_{q > p} V_s^q + \sum_{t \in s} \sum_{r \in S_{i(s)}^+} \sum_{q > p} (v_r^t)^q + \sum_{t \in s} \sum_{r \in S_{i(s)}^-} \sum_{\forall q} (v_r^t)^q \quad (18)$$

$$\tilde{V}_s^p = \sum_{\substack{q > 2 \\ (1)}} V_s^q + \sum_{\substack{t \in s \\ r \in S_{i(s)}^+ \\ q \geq p \\ (2)}} \sum \sum (v_r^t)^q + \sum_{\substack{t \in s \\ r \in S_{i(s)}^- \\ \forall q \\ (3)}} \sum \sum (v_r^t)^q \quad (19)$$

Where:

$(v_s^t)^p$: Flow in Tons, corresponding to cars with priority

$S_{i(s)}^+$: Set of route sections with common initial yard $i(s)$.

$S_{i(s)}^-$: Set of route sections, with initial yard before $i(s)$ and final yard after $i(s)$, that contain services also included in s .

The first term in equation (19) represents flows with higher priority, using the same route section s ; the second term represents same priority (competing) or higher priority flows, originated in the same yard $i(s)$ and using services contained in route section s , but corresponding to other route sections; finally, the third term represents flows with higher priority, that use services contained in route section s , but in route sections with origin before and destination after the yard $i(s)$.

Notice that car flows on section s , f_s^{vp} , used in equation (12) and tons flows V_s^{vp} present the following relation:

$$V_s^{vp} = \omega_v^\varepsilon \cdot f_s^{vp,\varepsilon} + \sum_{\pi} \omega_v^\pi \cdot f_s^{vp,\pi} \quad (20)$$

Marginal cost expression for the delays corresponding to each type of arc considered are developed in Appendix 1.

5. MATHEMATICAL FORMULATION

Now we can formulate the rail operator problem as the following variational inequality:

$$\mathcal{C}^T(\hat{X}^*) \cdot (\hat{X} - \hat{X}^*) - G^T(\hat{F}^{\varepsilon*}) \cdot (\hat{F}^\varepsilon - \hat{F}^{\varepsilon*}) \geq 0 \quad ; \forall \hat{X}, \hat{F}^\varepsilon \in \Omega \quad (21)$$

where, $\mathcal{C}^T = \{CCO_a^{vp}, CCI_a, CS_s^{vp}\}$ is the vector of arc cost functions corresponding to the auxiliary service network $G(N, A)$ defined in point 4. $\hat{X}^* = \{\hat{F}^*, \hat{V}^*\}$ is the vector of equilibrium arc flows in the same network, where $\hat{F}^* = \{f_a^{vp,\pi}\}$ is the vector of rail car flows, corresponding to consolidation and classification arcs and $\hat{V}^* = \{V_s^{vp,\pi}\}$ is the vector of equilibrium flows, corresponding to route sections expressed in Tons. $\hat{X} = \{\hat{F}, \hat{V}\}$ is any vector of feasible flows. $G^T = \{g_w^{vp,\varepsilon}\}$ is the vector containing inverse demand functions for empty cars and $\hat{F}^{\varepsilon*} = \{T_w^{vp,\varepsilon*}\}$ is the vector of equilibrium O-D demands for empty cars. Finally, $\hat{F}^\varepsilon = \{T_w^{vp,\varepsilon}\}$ represents any vector of feasible demands for empty cars. Ω is the set containing all feasible vectors of flows and empty cars demands.

In general, problem (21) does not have an equivalent optimization problem, because the Jacobian corresponding to the arc cost functions on the auxiliary service network $J(\mathcal{C}^T(\hat{V}^*))$ (specially because route section flow delay functions (15) and consolidation delay functions (11)), are non symmetric (Florian and Spiess, 1982).

The feasible set Ω is defined by the following linear constraints:

$$T_w^{vp,\varepsilon} = \sum_{r \in R_w} h_r^{vp,\varepsilon} \quad ; \forall w \in W, \forall v \in V, \forall p \quad (22)$$

$$T_w^{vp,\pi} = \sum_{r \in R_w} h_r^{vp,\pi} \quad ; \forall w \in W, \forall v \in V, \forall p, \forall \pi \quad (23)$$

$$O_i^{vp,\varepsilon} = \sum_{j \in D} T_w^{vp,\varepsilon} \quad ; \forall i \in O, \forall v \in V, \forall p \quad (24)$$

$$D_j^{vp,\varepsilon} = \sum_{i \in O} T_w^{vp,\varepsilon} \quad ; \forall j \in D, \forall v \in V, \forall p \quad (25)$$

$$f_a^{vp} = \sum_{r \in R} \delta_{ar} \cdot \left(\sum_{\pi} h_r^{vp,\pi} \right) \quad ; \forall a \in Co, \forall v \in V, \forall p \quad (26)$$

$$f_a = \sum_{r \in R} \delta_{ar} \left(\sum_v \sum_p \left(h_r^{vp,\varepsilon} + \sum_{\pi} h_r^{vp,\pi} \right) \right) \quad ; \forall a \in Cl \quad (27)$$

$$V_s^{vp} = \sum_{r \in R} \delta_{sr} \left(h_r^{vp,\varepsilon} \cdot \bar{\omega}_v^\varepsilon + \sum_{\pi} h_r^{vp,\pi} \cdot \bar{\omega}_v^\pi \right) \quad ; \forall s \in S, \forall v \in V, \forall p \quad (28)$$

$$V_s^{vp} = \bar{\omega}_v^\varepsilon \cdot f_s^{vp,\varepsilon} + \sum_{\pi} \bar{\omega}_v^\pi \cdot f_s^{vp,\pi} \quad ; \forall s \in S, \forall v \in V, \forall p \quad (29)$$

$$\delta_{ar} = \begin{cases} 1 & \text{If } a \in r \\ 0 & \text{If } a \notin r \end{cases} \quad ; \forall a \in \{Co \cup Cl\}, \forall r \in R \quad (30)$$

$$\delta_{sr} = \begin{cases} 1 & \text{If } s \in r \\ 0 & \text{If } s \notin r \end{cases} \quad ; \forall s \in S, \forall r \in R \quad (31)$$

$$T_w^{vp,\varepsilon} \geq 0 \quad ; \forall w \in W, \forall v \in V, \forall p \quad (32)$$

$$h_r^{vp,\varepsilon} \geq 0 \quad ; \forall w \in W, \forall v \in V, \forall p \quad (33)$$

$$h_r^{vp,\pi} \geq 0 \quad ; \forall w \in W, \forall v \in V, \forall p, \forall \pi \quad (34)$$

Conditions (22) and (23) are the classical flow conservation constraints expressed in terms of path flows; constraints (24) and (25) are the typical origin and destination constraints to make sure that the flows obtained are compatible with the demands. Constraints (26), (27) and (28) are necessary to make sure that arc flow and path flow solutions are consistent: the first one relates to flows on consolidation arcs, the second to flows on classification arcs and the last one to flows on route sections. In these, the flows expressed in cars must be transformed to Tons, because delays experienced in route sections are function of the Tons transported. Finally, constraints (30) and (31) correspond to arc-route incidence matrixes that define the network topology (for yard arcs and route sections) and constraints (32), (33) and (34) are typical non negativity constraints.

It is important to notice that a typical solution of the problem gives equilibrium flows for each route section of the auxiliary service network. In order to obtain equilibrium flows for each of the services or trains t , we will assume that flows are distributed among services proportional to their frequencies:

$$v_{t,s}^{vp} = \frac{freq_t \cdot V_s^{vp}}{freq_s} \quad ; \forall t \in s, \forall s \in SR, \forall v \in V, \forall p \quad (35)$$

$$\text{where, } freq_s = \sum_{t \in s} freq_t$$

The solution obtained using equation (35) is only a linear approximation to the exact solution. In order to obtain the last one it is necessary to solve a non-linear system of equations, that is obtained when effective frequencies (instead of nominal frequencies) are used (see De Cea and Fernández, 1993). However this last approach significantly complicates the solution of the problem making it very hard to solve. Because of this heuristics approaches have been proposed (see De Cea and Fernández, 1995) which have given excellent results in practice.

6. SOLUTION ALGORITHMS

Given that the vector cost function $\hat{C}^T(\hat{X})$ in (21), will in general have an asymmetric Jacobian, the problem formulated in section 5 will not have an equivalent optimization formulation. Therefore, a solution approach must use an algorithm that is able to solve directly variational inequalities (21), like the cutting plane algorithm (Nguyen and Dupuis, 1984) or some other methods to solve variational inequalities (Harker and Pang, 1987).

However, one approach that has been used successfully to solve asymmetric network assignment problems in practice, is the popularly known “diagonalization” method (Florian, 1977; Abdulaal and LeBlanc, 1979). The method corresponds to an iterative approach of a general Jacobi type for solving nonlinear equations (Pang and Chan, 1982). At each iteration, the vector cost function is “diagonalized” at the current solution, yielding a symmetric network problem.

In our case, at any given iteration of the diagonalization algorithm, a symmetric network problem of the following form is obtained:

$$\begin{aligned} \underset{\{F, V, T^\varepsilon\}}{\text{Min}} Z = & \sum_{a \in Co} \sum_v \sum_p \int_0^{f_a^{vp}} \hat{C} Co_a^{vp}(x) dx + \sum_{a \in Cl} \int_0^{f_a} C Cl_a(x) dx \\ & + \sum_{s \in SR} \sum_v \sum_p \int_0^{V_s^{vp}} \hat{C} SR_s^{vp}(x) dx \\ & + \frac{1}{\gamma} \sum_{w \in W} \sum_v \sum_p T_w^{vp, \varepsilon} \cdot (\ln T_w^{vp, \varepsilon} - 1) \end{aligned} \quad (36)$$

$$\text{s.a.: } \hat{F}, \hat{V}, \hat{T}^\varepsilon \in \Omega$$

Where hats, $\hat{}$, over the consolidation arcs and route sections mean that those variables are diagonalized.

A sequence of such problems must be solved until a sufficiently stable solution is obtained. The most popular algorithm for solving problem (36) is the well known Frank-Wolf descent method (LeBlanc et al, 1975). Convergence of the algorithm is based on the usual monotonicity condition required for the arc cost function in (36) (Florian and Spiess, 1982).

7. CONCLUSIONS

We have proposed a new model to represent the behavior of freight rail network systems. The model presents several new characteristics that provide more realism to the analysis performed: i) Railway services and their operational characteristics are explicitly considered, ii) The distribution of empty cars and its assignment over the rail network is treated simultaneously with the assignment of products to be transported, considering the operating costs over the network and the transportation priorities, iii) Different priorities are assigned to different types of cars to be moved depending on the products transported, iv) Capacity constraints are considered for the movement of different products, depending on the availability of the type of cars necessary and the demands of products competing for the use of the same cars. In a next work we analyze different methods to solve variational inequality (21) directly and we compare them with the performance of the diagonalization algorithm proposed above.

ACKNOWLEDGMENTS

This research was supported by FONDECYT (Chile) and the Universidad Católica de Chile.

REFERENCES

- Abdulaal, M. and LeBlanc, L.J. (1979) Methods for Combining Modal Split and Equilibrium Assignment Models. *Tptn. Res.*, 13, 292-314.
- Assad, A.A.(1980) Models for Rail Transportation. *Tptn. Res.*, 14A, 205-20.
- Bouley, J. (1987) Just in Time. *Railway Gaz. Int.*, 2, 93-95.
- 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. MSc. thesis, Publicación 118, Centre de Recherche sur les Transports, Université de Montréal.
- Crainic, T.G., Ferland, J.A. and Rousseau, J.M. (1984) A Tactical Planning Model for Rail Freight Transportation. *Tptn. Sc.*, 18(2), 165-184.
- Crainic, T.G., Florian, M. and Guélat, J. (1990) A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows. *Tptn. Sc.* 24(1), 25-39.
- Crainic, T.G., Florian, M. and Léal, J.E. (1990) A Model for the Strategic Planning of National Freight Transportation by Rail. *Tptn. Sc.*, 24(1), 1-24.
- Crainic, T.G. and Gendreau, M. (1986) Approximate Formulas for the Computation of Connection Delays under Capacity Restriction in Rail Freight Transportation. *Proceeding of the fourth World Conference on Transport Research*, May 1986, Vancouver, Canada.
- De Cea, J. and Fernández, J.E.(1993) Transit Assignment for Congested Public Transport Systems: An Equilibrium Model. *Tptn. Sc.*, 27(2), 133-147.
- Dejax, P. and Crainic, T.G. (1987) A Review of Empty Flows and Fleet Management Models in Freight Transportation. *Tptn. Sc.*, 21(4), 227-247.
- English, G.W. (1977) An Analytic Model for the Analysis of Single Track Railway Capacity. Presented at International Symposium of Travel Supply Models, Centre de Recherche sur les Transports, Université de Montréal.
- Fernández, J.E., J. De Cea and A. Soto (1998) Multimodal Equilibrium Models for Predicting Intercity Freight Flows. Presented to *Tptn. Res.* for publication.
- Florian, M. (1977) A Traffic Equilibrium Model of Travel by Car and Public Transit Modes. *Tptn. Sc.*, 8, 166-179.
- Florian, M. and Spiess, H. (1982) The Convergence of Diagonalization Algorithm for Asymmetric Network Equilibrium Problems. *Tptn. Res.*, 16B(2), 447-483.

- Folk, J.F. (1972) Some Analysis of Railroad Data. Studies in Railroad Operation and Economics, Vol. 6, Report No. R72-41, Department of Civil Eng., MIT, Cambridge, MA.
- Franck, O. (1966) The Two Way Traffic on a Single Line of Railway. Operation Research, 14, 801-811.
- Friesz, T.L., Gottfried, J.A. and Morlok E.K. (1986) A Sequential Shipper-Carrier Network Model for Predicting Freight Flows. Tptn. Sc., 20(2), 80-91.
- Friesz, T.L., Tobin, R.L. and Harker, P.T. (1981) Variational Inequalities and Convergence of Diagonalization Methods for Derived Demand Network Equilibrium Problems. Report CUE-FNEM-1981-1-1, Dept. of Civil Engineering, University of Pennsylvania, October.
- Harker, P.T. and Pang, J. (1987) Finite-Dimensional Variational Inequalities and Nonlinear Complementarity Problems: A Survey of Theory, Algorithms and Applications. Department of Decision Sciences, the Wharton School, University of Pennsylvania.
- Hautzinger, H. (1984) The Prediction of Interregional Goods Vehicle Flow: Some New Modeling Concepts. 9th Int. Symposium on Transportation and Traffic Theory, pp. 375-396, J. Vormuller and R. Hamerslag (eds.), VNU Science Press, The Netherlands, 1984.
- Hay, W. (1982) Rail Road Engineering. John Wiley and Sons, New York, second edition.
- Jovanovic, D. and Harker, P.T. (1991) Tactical Scheduling of Rail Operations: The SCAN I System. Tptn. Sc., 25(1), 46-64.
- LeBlanc, L.J., Morlok, E.K. and Pierskalla, W.P. (1975) An Efficient Approach to Solving the Road Network Traffic Assignment Problem. Tptn. Res., 5, 309-318.
- Marín, A., Menéndez, A.L. and Salmerón, J. (1993) Planificación Táctica del Transporte de Mercancías por Ferrocarril, Estudios de Transportes y Comunicaciones, N°61, Oct-Dic 1993, Ministerio de Obras Públicas, Transportes y Medio Ambiente de España.
- Mendiratta, V.B. (1982) A Dynamic Optimization Model of the Empty Car Distribution Process. Ph.D. Thesis, Department of Civil Eng., Northwestern University.
- Nguyen, S. and Dupuis, C. (1984) An Efficient Method for Computing Traffic Equilibria in Networks with Asymmetric Transportation Costs, Tptn. Sc., 18, 185-202.
- Pang, J.M. and Chan, D. (1982) Iterative Methods for Variational and Complementarity Problems. Math. Program, 24, 284-313.
- Payraud, R. (1981) La Modélisation du Transport des Marchandises: Le Modèle MASUP. Revue Générale des Chemin de Fer, 743-750.
- Petersen, E.R. (1974), Over-the-Road Transit Time for Single Track Railway, Tptn. Sc., 8, 65-74.
- Petersen, E.R. (1977a), Railyard Modeling: Part I: Prediction of put-through Time, Tptn. Sc., 11(1), 37-49.
- Petersen, E.R. (1977b), Railyard Modeling: Part II: The Effect of Yard Facilities on Congestion, Tptn. Sc., 11(1), 50-59.
- Petersen, E.R. (1977c) Capacity of a Single Track Rail Line. School of Business Working Paper No.77-38, Queen's University, Kingston, Ontario, Canada.
- Petersen, E.R. and Fullerton, H.V. (1975) The Railcar Network Models. Canadian Institute of Guided Ground Transport, Queen's University, Kingston, Ontario, Canada.
- Petersen, E.R. and Taylor, A.J. (1982), A Structured Model for Rail Line Simulation and Optimization, Tptn. Sc., 16(2), 192-205.
- Reebie Associates (1972) Toward an Effective Demurrage System. prepared for the US Department of Transportation, Federal Railroad Administration, 1972.
- Reid, R.M., O'Doherty, J.D., Sussman, J.M. and Lang, A.S. (1972) The Impact of Classification Yard Performance on Rail Trip Time Reliability, Studies in Railroad Operations and economics 4, Department of Civil Eng., MIT, Cambridge, MA.

Thomet, M.A. (1971) A Combinatorial Search Approach to the Freight Scheduling Problem. Ph.D. Thesis, Department of Electrical Eng., Carnegie Mellon University.
 Turnquist, M.A. and Daskin, M.S. (1982), Queuing Models of Classification and Connection Delay in Railroads, Tptn. Sc., 16(2), 207-230.
 Wardrop, J.G. (1952), Some Theoretical Aspects of Road Traffic Research, in Proceedings of the Institute of Civil Engineering, Part II, pp. 325-378

APPENDIX : Marginal Delay Functions

Here we present the marginal delay functions for each arcs of the rail network developed in section 4. This functions are calculated using the following relation:

$$DM\bar{g}_a(f_a) = \frac{\partial}{\partial f_a}(f_a \cdot DM\bar{e}_a(f_a)) = \underbrace{DM\bar{e}_a(f_a)}_{(1)} + f_a \cdot \underbrace{\frac{\partial DM\bar{e}_a(f_a)}{\partial f_a}}_{(2)} \quad (1.1)$$

Taking into account that the first term (1) in equation (1.1) was calculated for each arc type in section 4, in this appendix we only present the determination of the third term of this equation, for each arc type of the network.

Consolidation Marginal Delays

In the case of consolidation delays, taking into account that average delay in this operation depends on the flows of classification and route sections arcs (see equation (11)), but is not dependent of its own flow, then the diagonalized average delay function on the consolidation arcs is constant¹. Therefore, in this case the third term of equation (1.1) is equal to zero and the average delay is equal to the marginal delay.

Classification Marginal Delays

To calculate classification marginal delays we consider the following equation presented in section 4.

$$DMeCl_a = DFCl_a + \beta Cl_a \left(\frac{f_a}{CAP_a} \right)^{nCl_a} \quad (1.2)$$

Notice that classification arcs delays (see equation (1.2)) depend only on there own flows. The third term of equation (1.1) is then equal to:

$$\frac{\partial DMeCl_a(f_a)}{\partial f_a} = \frac{\beta Cl_a \cdot nCl_a}{CAP_a} \left(\frac{f_a}{CAP_a} \right)^{nCl_a-1} \quad (1.3)$$

Connection and Route Sections Marginal Delays

To calculate the third term of equation (1.1) we use the following diagonalized route section average delay function.

$$DM\bar{e}_s^p(V_s^p) = DFS_s + \left(\frac{\alpha}{f_s} \right) + \beta S_s \cdot \left(\frac{V_s^p + \bar{V}_s^p}{K_s} \right)^{nS_s} \quad (1.4)$$

Therefore, following a similar approach to that used before, the third term of the marginal delay on a route section is given by:

$$\frac{\partial DM\bar{e}_s^p(V_s^p)}{\partial V_s^p} = \frac{\beta S_s \cdot nS_s}{K_s} \cdot \left(\frac{V_s^p + \bar{V}_s^p}{K_s} \right)^{nS_s-1} \quad (1.5)$$

¹ Notice that shipment wait in consolidation arcs until an appropriate type of arc is made available; then the shipment is loaded an the cars leave the arc. We consider that required for the car loading operations can be negligible and therefore consolidation arc do not depend on car flows.