A Modeling Framework for Passenger Assignment on a Transport Network with Timetables

Sang Nguyen • Stefano Pallottino • Federico Malucelli
Département d'informatique et de recherche opérationnelle and Centre de recherche sur les transports, Université de Montréal, c.p. 6128, succ. centre ville, Montréal, Canada H3C 3J7
Dipartimento di Informatica, Università di Pisa, 40 corso Italia, 56125 Pisa, Italy
Dipartimento di Elettronica e Informazione, Politecnico di Milano, P. le L. da Vinci 32, Milano, Italy

This paper presents a new graph theoretic framework for the passenger assignment problem that encompasses simultaneously the departure time and the route choice. The implicit FIFO access to transit lines is taken into account by the concept of available capacity. This notion of flow priority has not been considered explicitly in previous models. A traffic equilibrium model is described and a computational procedure based on asymmetric boarding penalty functions is suggested.

Introduction

Consider a network of buses and/or surface and underground railway lines, such that every transit line (bus or railway) has a fixed itinerary and is described by a sequence of bus stops or rail stations and a fixed set of scheduled trips. We are interested in the case where the time-tables are reasonably reliable, and the number of transit lines and the trip frequency of these lines are lower than that commonly associated with urban networks. For such networks—usually associated with the suburban or inter-urban setting—the departure time and the route choice are equally important to commuters. Previous studies that dealt with the departure time decision problem concentrated mainly on the deterministic and stochastic highway traffic assignment models. In addition, most of the existing models are limited to a single-route or single origin-destination pair (corridor network), or to either the departure time choice or to the route selection but not to both dimensions simultaneously (e.g. Mahmassani and Herman 1984, Alfa 1986, Mahmassani and Chang 1986, 1987, Newell 1987, Carey and Srinivasan 1988, Friesz et al. l989, Alfa 1989). Only a few researchers have addressed the departure time and route choice decision problem for public transportation network (Hendrickson and Plank 1984, Sumi et al. 1990). However, no general framework for practical size networks has been developed to date, and existing frameworks for urban networks, such as the hyperpath graph-theoretic framework (e.g. Gendreau l984, Nguyen and Pallottino 1988, Spiess and Florian 1989, Wu et al. 1994), are not directly applicable since they are intrinsically static and do not take into account the time-tables.

We consider in this paper a new graph-theoretic framework for the passenger assignment problem that encompasses simultaneously the departure time and the route choice dimensions. The proposed framework is built upon the central concept of the *path available capacity* that allows us to capture the flow priority aspect. This latter, induced by the implicit FIFO rule that binds passengers at every access point to the transit network, has not been taken into account explicitly in previous models, although similar properties were analyzed in Carey (1992). Various illustrative examples highlight the distinctive characteristics of the problem considered. Finally we

0041-1655/01/3503/0238$05.00
1526-5447 electronic ISSN

provide a formulation of the user-optimal assignment problem and suggest a computational procedure for determining an equilibrium flow based on asymmetric boarding penalty cost functions. A distinct direction of research based on the concepts of *strategic equilibrium flow* is investigated in Marcotte and Nguyen (1998).

1. The Public Transportation Network and the Space-Time Graph

We will first describe the basic transportation network. Consider a set of transit lines $\mathcal{L} = \{L_1, L_2, \ldots, L_l\}$, where every line is defined by a sequence of trips which cover the same itinerary in different scheduled times. Each itinerary is a sequence of geographical locations—bus stops or railway stations—referred to in the sequel as *transit nodes*. To each trip is associated a carrier with a given capacity and a reasonably reliable timetable, i.e., a pair of arrival and departure times for every transit node of the itinerary. Origins and destinations are connected to the transit nodes by walking links.

To be able to mathematically describe the commuters' departure times and itinerary choices on this network, we introduce the following *space-time graph* $G = (O \cup D \cup N, A)$. O and D are respectively the set of time-invariant origins and destinations, N is the set of space-time nodes, and A is the set of space-time arcs.

Each node $i \in N$ represents a transit node at a particular time $t(i)$. An arc $(i, j) \in A$ describes a movement departing the *tail* i at time $t(i)$ and arriving at the *head* j at time $t(j)$, and $t(j) \geq t(i)$. Arcs may be divided into two categories. The first group includes:

- *in-vehicle arcs* representing the portion of a trip departing from i at time $t(i)$ and arriving at j at time $t(j)$;
- *access arcs*, where i is an origin and j is a transit node in the neighborhood of i;
- *egress arcs*, where i is a transit node and j is a destination in the neighborhood of i;
- *walking arcs*, where i and j are two neighboring transit nodes.

The second group includes arcs that represent movements in time only (i.e., i and j represent the same transit node), such as:

- *boarding arcs*, where i is the head node of an access or a walking arc and j is the tail node of an in-vehicle arc;
- *leaving arcs*, where i is the head node of an in-vehicle arc and j is the tail node of an egress or a walking arc;
- *stationary arcs*, where i and j are respectively the head and the tail nodes of in-vehicle arcs representing the same trip;
- *transfer arcs*, where i and j are respectively the head and the tail nodes of in-vehicle arcs representing different trips.

To every access, egress, and walking arc (i, j) is associated a walking time $\lambda(i, j)$. Each transit node s is described by a bipartite subgraph $B(s) = (H_s, K_s, A_s)$ of the space-time graph G, where $H_s = \{i \mid i$ is the head node of an access arc, or an in-vehicle arc, or a walking arc$\}$, $K_s = \{j \mid j$ is the tail node of an in-vehicle arc, or an egress arc, or a walking arc$\}$, and $A_s = \{(i, j) \mid i \in H_s, j \in K_s, (i, j)$ is a boarding arc, or a leaving arc, or a stationary arc, or a transfer arc$\}$ (see Figure 1).

A trip is described on this space-time graph by an alternating sequence of stationary and in-vehicle arcs, and a passenger's route from an origin o to a destination d is described by a path $\{o, i_1, \ldots, i_k, d\}$ starting at o at time $(t(i_1) - \lambda(o, i_1))$ with an access arc (o, i_1)

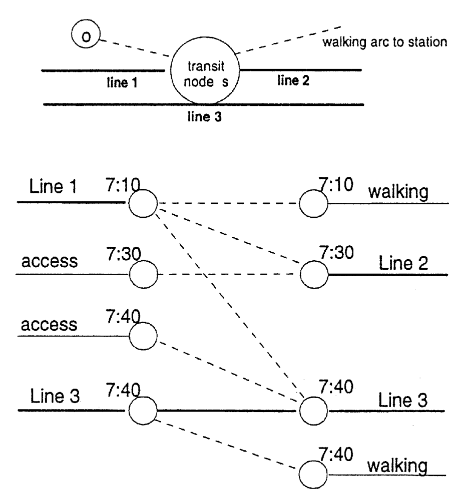

Figure 1 Bipartite Subgraph Associated with a Transit Node

and reaching d at time $(t(i_k)+\lambda(i_k, d))$ with an egress arc (i_k, d). The subpath $\{i_2, \dots, i_{k-1}\}$ is partitioned into trip segments connected together, either by a transfer arc, or by a sequence of leaving, walking, and boarding arcs. Note that the space-time graph introduced is *acyclic*, since arcs (i, j) with $t(i)=t(j)$ are all included in various acyclic bipartite subgraphs and origins are separated from destinations.

To explicitly capture the implicit FIFO rule that binds passengers at every transit node s, we consider in the sequel that the set of arcs incident into each node $j \in K_s$—the backward star $S^-(j)$ of node j (set of arcs incident into j)—is an ordered set

$$S^-(j)=\{(i_0, j)\}\cup\{(i_1, j), (i_2, j), \dots, (i_n, j)\}, \quad (1)$$

where (i_0, j) is a stationary arc, if such an arc exists, $n=|S^-(j)|$, and

$$t(i_1)\le t(i_2)\le \cdots \le t(i_n). \quad (2)$$

Passengers arriving with the carrier and staying for the next leg correspond to the flow on the stationary arc, while the other passengers share the residual capacity of the carrier in the order of their respective arrival times at transit node s.

Consider the morning peak period. Passengers are subdivided into subgroups indexed by g. Passengers of the same group g share the same *desired arrival time interval* $[t^-(g), t^+(g)]$ at their respective destinations. In what follows, the demand for any given origin-destination pair (o, d) and subgroup g is denoted by $\mathcal{F}(o, d, g)$.

Each passenger must decide when to leave his origin and which transportation service to use to reach his destination ideally within the desired time interval.

2. Capacity and Flow Priority

In contrast to the road network case, where it is usually considered that the cost of using an arc is an increasing function of traffic flow, the number of passengers aboard a vehicle of a public transport network does not have a direct influence on the length of a trip, which is assumed to be constant. Nevertheless, congestion may force a passenger to take a lengthier path to destination because the vehicle that he plans to board may not have any available space. It can be seen that an inherent characteristic is the asymmetric aspect of passengers inter-influence, and this must be explicitly dealt with. We first focus on the approach with explicit capacity constraints and then exploit the possibility of approximating these capacity constraints with appropriate penalty cost functions. We also adhere to the continuous approximation of passenger flow for practical reasons, although a discrete formulation may seem more suitable for small networks.

Let h_p^g denote the portion of the demand $\mathcal{F}(o, d, g)$ traveling on path $p\in P_{od}$, where P_{od} is the set of all paths connecting the pair (o, d). The usual equations of conservation of flow are

$$\sum_{p\in P_{od}} h_p^g=\mathcal{F}(o, d, g), \quad \forall (o, d, g),$$
$$h_p^g\ge 0, \quad \forall p\in P_{od}, \forall (o, d, g). \quad (3)$$

The following inequalities express the usual capacity constraints:

$$\sum_{(o,d,g)}\sum_{p\in P_{od}} \delta_{ap} h_p^g \le u_a, \quad \forall a\in A_v, \quad (4)$$

where A_v is the set of in-vehicle arcs or stationary arcs, u_a is the capacity of the vehicle associated with arc a, and $\delta_{ap}=1$, if path p traverses arc a and 0 otherwise. A *feasible flow* is a vector $\mathbf{h}=\{h_p^g\}$ satisfying the set of constraints (3), and a feasible flow is called *compatible* if it also satisfies equation (4). Let Λ denoted the set of feasible flows $\mathbf{h}$, and let $\Omega\subseteq\Lambda$ be the set of compatible flows. It will be assumed that Ω is nonempty. Let x_a^{gd} denote the partial flow associated with the pair (g, d) on arc $a\in A$,

$$x_a^{gd}=\sum_{o\in O}\sum_{p\in P_{od}} \delta_{ap} h_p^g, \quad \forall a\in A. \quad (5)$$

The total flow y_a on arc a is then

$$y_a=\sum_{g}\sum_{d\in D} x_a^{gd}. \quad (6)$$

$\mathbf{x} = \{x_a^{gd}\}$ and $\mathbf{y} = \{y_a\}$ are feasible if $\mathbf{h}$ is feasible and compatible whenever $\mathbf{h}$ is itself compatible.

2.1. Generalized Arc Costs and Path Costs

To model the passenger's choice of a departure time and a route, we must define a path disutility cost that, in addition to the usual perceived path travel time, also includes penalty costs associated with an early departure as well as early or late arrival. To design the early and late departure costs, we will first determine a *free-flow latest-departure-time* $\tau(o, d, g)$ that represents the ideal time to leave origin o to get to destination d within the time interval $[t^-(g), t^+(g)]$, for each triplet (o, d, g).

The departure time $\tau(o, d, g)$ is determined by identifying the subset of paths $P_{odg} \subseteq P_{od}$ of the space-time graph that reach destination d within the time interval indexed by g. This can be achieved with a backward visit of the time-space graph starting from the set of nodes

$$Q := \{i \in N \mid (i, d) \in S^-(d);\\ t^-(g) \le t(i) + \lambda(i, d) \le t^+(g)\},$$

assumed nonempty followed by the scanning of the forward star $S^+(o)$ (set of arcs incident out of o) of every origin $o \in O$. The computation of $\tau(o, d, g)$ for a given pair (d, g) may be carried out with the following simple connectivity checking procedure on the space-time graph:

Procedure. *Latest_departure_time* $(g, d, t^-(g), t^+(g), \boldsymbol{\tau})$;
begin
  $Q := \{i \in N \mid (i, d) \in S^-(d),\ t^-(g) \le t(i) + \lambda(i, d) \le t^+(g)\}$;
  for $i \in N$, $label[i] := 0$;
  for $i \in Q$, $label[i] := 1$; **for** $o \in O$, $label[o] = 1$;
  repeat
    fetch $u \in Q$;
    $Q := Q - \{u\}$;
    for each $(i, u) \in S^-(u)$ such that $label[i] = 0$
    begin
      insert i into Q;
      $label[i] := 1$
    end
  until $Q = \varnothing$;
  for $o \in O$, $\tau(o, d, g) = \max\{t(u) - \lambda(o, u) \mid (o, u) \in S^+(o),\ label[u] = 1\}$
end.

Once $\tau(o, d, g)$ is determined, we can define a penalty cost for every access arc (o, j) which depends on the difference between the actual departure time $t(j) - \lambda(o, j)$ and the latest-departure-time. For instance

$$U_{oj}^{odg} = \max\{\mu(\tau(o, d, g) - t(j) + \lambda(o, j)), 0\}, \tag{7}$$

where μ is a positive scalar. Note that whenever $\tau(o, d, g) < t(j) - \lambda(o, j)$, then no penalty is incurred at the origin although the arrival time at the destination does not belong to the desired time interval.

Similarly, late and early arrival at destination also induces a penalty cost (see, for instance, De Palma et al. 1983, Friesz et al. 1993). Thus, for every egress arc (i, d) we have

$$U_{id}^g = \begin{cases} \eta_1(t^-(g) - t(i) - \lambda(i, d)), & \text{if } t(i) + \lambda(i, d) < t^-(g), \\ \eta_2(t(i) + \lambda(i, d) - t^+(g)), & \text{if } t(i) + \lambda(i, d) > t^+(g), \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $\eta_2 \ge \eta_1 > 0$ are user defined scalars. At this point, the cost C_p^g of any given path $p = \{o, i_1, \ldots, i_k, d\} \in P_{od}$ for passengers of a group g, may be defined unambiguously as the sum of the lengths of its arcs and the penalty terms associated with the access and the egress arc

$$C_p^g = \sum_{(i,j)\in p} l(i, j) + U_p^g,$$

where $l(i, j)$ is a generalized cost given to each arc (i, j) to differentiate the perceived value of time of the various elements composing a trip, and

$$U_p^g = U_{oi_1}^{odg} + U_{i_k d}^g. \tag{9}$$

It should be noted here that the latest-departure-time *path* itself is not relevant at this stage. Only the latest-departure-times are needed for calibrating the penalty costs U_{oj}^{odg}. However, it is worth mentioning that travel times in transit networks, in general, do not satisfy the first-in-first-out property and thus later

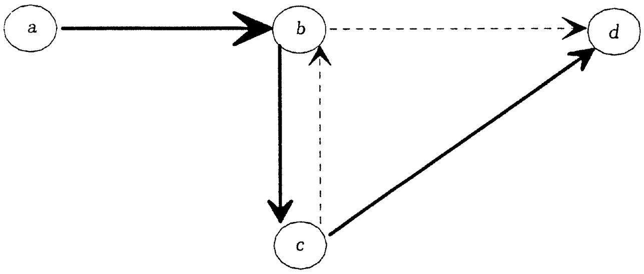

Figure 2 **Path with Cycle**

departure may arrive earlier (i.e., an express line versus a regular one). This may result in generating paths that may contain one or several cycles. Such paths are not always unrealistic. For instance, consider the example given in Figure 2, with the three lines:

$$l_1=\{(a,b),(b,c)\},\quad l_2=\{(c,b),(b,d)\},\quad l_3=\{(c,d)\}.$$

A passenger going from a to d may follow one of the three paths:

1. line l_1 to b then l_2 to d,
2. line l_1 to b then c, and l_3 to d,
3. or line l_1 to b then c, and l_2 back to b then d.

The last path goes through node b twice. Such a path is perfectly reasonable in many cases. For instance, assume that l_2 is a congested line and the risk of being unable to board it at b is much higher than at c. Then making a detour to node c to ensure a place on the carrier of l_2 would produce the above cycle. Similarly, if line l_3 is a congested express line, then a passenger may accept the risk of making an unsuccessful detour to c to board the carrier of l_3. The boarding penalty costs, introduced in §3.1, may indeed allow generating such paths.

2.2. Available Capacity and Flow Priority

One may interpret the assignment of passengers on a public transport network as an n-person noncooperative game, and a Wardrop (1952) equilibrium flow as a flow such that no individual can improve his travel cost by unilaterally switching to another feasible path considering the other passenger's choices as fixed. With the FIFO rule at the transit nodes, path feasibility cannot be defined solely in terms of the residual capacity of the carriers. For instance, consider the example given in Figure 3.

In Figure 3, assume that the capacity of every carrier is 20, and a single group of 21 passengers going from o to d with the desired arrival time interval $[9h00, 9h15]$. Access and egress arcs are assumed to have zero length. With $\mu=1$, the latest departure time $\tau=8h45$ induces penalty costs $U_{oc}=30$, $U_{ob}=15$, and $U_{oa}=0$ on the access arcs, and $U_{kd}=U_{ld}=0$ on the egress arcs. The shortest path is $p_1=\{o,a,e,h,j,k,d\}$, with $C_{p_1}=30$. Let $p_2=\{o,b,f,i,j,k,d\}$ and $p_3=\{o,c,g,l,d\}$ with $C_{p_2}=60$ and $C_{p_3}=75$, respectively. If feasibility is defined with respect to the usual arc residual capacities, then the flow $\{h_{p_1},h_{p_2},h_{p_3}\}=\{20,0,1\}$ satisfies the above definition of equilibrium, since arc (j,k) is saturated and thus the passenger on p_3 cannot improve his current traveling cost by moving either to p_2 or p_1. However, any seat occupied at stop j by a passenger transferring from line 1 is *available* to passengers boarding line 2 at the previous stop f (FIFO rule). Consequently, a passenger on p_3 may improve his cost by switching to path p_2 and prevent one passenger from p_1 to board at stop j.

This leads to the introduction of the concept of *available capacity* to capture the priority order that

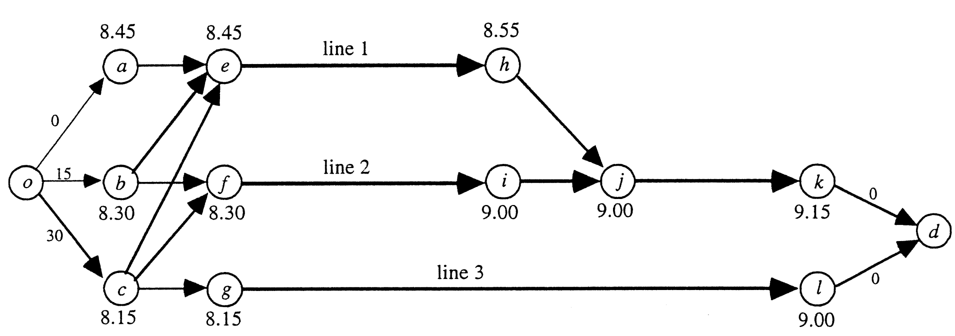

Figure 3 **Priority and Equilibrium Flow**

exists between the passengers already in the transit carrier and those trying to board it, on the one hand, and between passengers entering the boarding queue at different times, on the other hand. Let $a = (i, j)$ be an in-vehicle or a stationary arc, for a given flow $\mathbf{h}$; the *residual capacity* of $a \in A$ is defined as usual as

$$r_a = u_a - y_a.$$

Consider the bipartite subgraph $B(s) = (H_s, K_s, A_s)$ corresponding to a given transit node s, and a node $j \in K_s$, that is, the tail of an in-vehicle arc a of a given trip. Recall that for such a node, the set of arcs incident into j is identified by the ordered set $S^-(j)$ (Equations (1) and (2)). The available capacity for every boarding or transfer arc (i_m, j) of $S^-(j)$ is defined as

$$q_{i_m j} = u_a - \sum_{k=0}^{m} y_{i_k j}, \quad m = 1, \ldots, n, \tag{10}$$

where $n = |S^-(j)|$ and flow $y_{i_0 j} = 0$ if the stationary arc (i_0, j) does not exist (j is the beginning of a trip). This set of available capacities represents the sharing of the carrier's residual capacity among passengers according to their priorities at node j.

Let $A_v(p)$ denote the set of in-vehicle and stationary arcs of p, and $A_b(p)$ that of boarding and transfer arcs. The *residual capacity* R_p of p is

$$R_p = \min\{r_a \mid a \in A_v(p)\}, \tag{11}$$

and the *available capacity* Q_p of p is

$$Q_p = \min\{q_a \mid a \in A_b(p)\}. \tag{12}$$

It can be seen that

$$R_p \le Q_p, \quad \forall p \in P_{od}, \forall (o, d).$$

Indeed, consider a transfer or boarding arc $(i_m, j) \in A_b(p)$ of p and let $(j, l) \in A_v(p)$ denote the in-vehicle arc that immediately follows (i_m, j) on p. The residual capacity r_{jl} of (j, l) is:

$$r_{jl} = u_{jl} - y_{jl} = u_{jl} - \sum_{k=0}^{n} y_{i_k j} \le u_{jl} - \sum_{k=0}^{m} y_{i_k j} = q_{i_m j},$$

and thus

$$R_p \le \min\{q_a \mid a \in A_b(p)\} = Q_p.$$

Consequently, a saturated path p (with zero residual capacity) may still have a positive available capacity. For the example in Figure 3, it can be seen that with the flow $\{h_{p1}, h_{p2}, h_{p3}\} = \{20, 0, 1\}$, the capacities of the respective paths are

capacity	p_1	p_2	p_3
residual R_{p_i}	0	0	19
available Q_{p_i}	0	20	19

hence, a new passenger on path p_2 would force a user on path p_1 to switch to either p_2 or p_3.

3. Equilibrium Flow and Pitfalls

We now have all the ingredients needed to analyze the traditional equilibrium model. First, as discussed, it is necessary to redefine the path feasibility as follows: *a path is feasible if and only if its available capacity is positive*. With this in mind, the following standard definition of an equilibrium flow may be stated.

Definition. A compatible flow is an *equilibrium flow* if no passenger can improve his traveling cost by unilaterally switching to another *feasible path*, while the choices of all other passengers are fixed.

It is worth noting here that the above is the standard static definition of equilibrium since the path feasibility is defined with respect to the fixed current flow, thus any possible forced change due to a unilateral path switching is ignored. Similar considerations in terms of the path costs are discussed in Smith (1979).

Assume that a passenger is considering unilaterally switching from path p to path q and the current available capacity of q is zero and all the saturated boarding or transfer arcs of q belong to p as well. In this case, it is important to distinguish two modelling options:

1. q remains infeasible for this passenger even if he contributes to the flow of the saturated arcs,

2. q is feasible for this particular passenger since the arcs of q that are not in p have positive available capacity.

We refer to the first option as the *additional passenger option* and the latter as the *switching passenger* one. Going back to the example in Figure 3, the two modeling options produce drastically different results. With the additional passenger model, it can be seen that flow $\{h_{p_1}, h_{p_2}, h_{p_3}\} = \{0, 20, 1\}$ satisfies the above definition of equilibrium as the minimum cost path p_1 is infeasible for any *additional passenger*. Note that in contrast to the Wardrop equilibrium in ordinary vehicle networks, the two utilized paths are not the minimum cost paths here.

With the switching passenger model the existence of an equilibrium flow may not be warranted. Indeed, in the same example, any feasible flow that does not saturate arc (j, k) is trivially not an equilibrium flow and any feasible flow $\{h_{p_1}, h_{p_2}, h_{p_3}\} = \{20-x, x, 1\}$ is also not an equilibrium flow, since whenever $x > 0$, p_1 is always feasible for any passenger on p_2, and when $x = 0$, the passenger on p_3 can switch to p_2. Consequently there is no equilibrium flow.

It seems therefore that the additional passenger model is well suited for networks with priority and hard capacities. Let $Q_p(\mathbf{h})$ denote the available capacity of path $p \in P_{od}$ when the path flow is $\mathbf{h}$. Also define

$$K(g, p) = \{p' \mid p' \in P_{od} \text{ and } C_{p'}^g < C_p^g\},$$

as the set of paths that dominate path p, and

$$\widetilde{Q}_p(\mathbf{h}) = \begin{cases} \sum_{p' \in K(g,p)} Q_{p'}(\mathbf{h}), & \text{if } K(g,p) \neq \varnothing, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

In terms of the pseudo-costs $\widetilde{Q}_p(\mathbf{h})$, we may again state the standard Wardrop's equilibrium conditions as: a compatible flow $\mathbf{h}^* \in \Omega$ is an equilibrium flow if it satisfies

$$h_p^{g*} = 0 \quad \text{if } \widetilde{Q}_p(\mathbf{h}^*) > 0, \forall p \in P_{od}, \forall g. \tag{14}$$

In the next subsection, we propose a variational inequality formulation based on *arc penalty cost functions* that simultaneously handle both the capacity and the flow priority characteristics. This approach leads to a more practical computational procedure that obviates the explicit enumeration of paths connecting each origin-destination pair required by the above complementarity problem.

3.1. A Penalty Approach

Consider the set $S^-(j)$ (Equations (1) and (2)) of arcs incident into the tail node j of an in-vehicle arc (j, k). The objective here is to prevent individuals from taking infeasible paths in terms of available capacities. Let us associate a penalty cost $W_a(\mathbf{y})$ with every boarding or transfer arc $a = (i_m, j) \in S^-(j)$, $m > 0$. For example, the following standard functional form may be used

$$W_a(\mathbf{y}) = \left\{\alpha_a\left(\left[\sum_{k=0}^{m} y_{i_k j} - \rho_a u_a\right]^+\right)^\theta, \right. \quad a = (i_m, j) \mid m = 1, 2, \dots, n, \tag{15}$$

where α_a, ρ_a, and θ are positive scalars, and $[z]^+ = \max\{0, z\}$. For $2 \le \theta < \infty$, $W_a(\mathbf{y})$ is continuously differentiable. Therefore, in addition to the usual waiting time $l(i_m, j) = t(j) - t(i_m)$, a passenger also incurs an asymmetric boarding penalty cost. The designed asymmetry reflects the fact that the flow $y_{i_{m-1}j}$ contributes to the cost incurred by $y_{i_m j}$ while there is no reciprocal impact and this is clearly the desired objective. Note also that the penalty costs associated with the boarding arcs implicitly take care of the capacity constraint (4). The cost of a path $p \in P_{od}$, $C_p^g(\mathbf{y})$ (or $C_p^g(\mathbf{h})$, since $\mathbf{y}$ is a function of $\mathbf{h}$), incurred by a traveller of group g becomes

$$C_p^g(\mathbf{h}) = C_p^g(\mathbf{y}) = \sum_{(i,j)\in p} l(i,j) + U_p^g + W_p(\mathbf{y}), \tag{16}$$

where the arrival and departure penalty cost U_p^g is defined in Equation (9) and the boarding penalty cost $W_p(\mathbf{y})$ is

$$W_p(\mathbf{y}) = \sum_{a \in A_b(p)} W_a(\mathbf{y}). \tag{17}$$

With the proposed cost structure, the determination of an equilibrium flow $\mathbf{h}^* = \{h_p^{g^*}\}$ may now be formulated as a variational inequality problem (Smith 1979, Dafermos 1980):

$$\text{Find } \mathbf{h}^* \in \Lambda \text{ satisfying } (\mathbf{h}^* - \mathbf{h})^T C(\mathbf{h}^*) \le 0, \quad \forall \mathbf{h} \in \Lambda, \tag{18}$$

where $C(\mathbf{h}^*) = \{C_p^g(\mathbf{h}^*)\}$. The existence of a solution $\mathbf{h}^*$ is ensured since Λ is compact and convex, and

$C(\mathbf{h})$ is continuous by construction (Kinderlehrer and Stampacchia 1980). Because of the asymmetry mentioned above, $C(\mathbf{h})$ is not a gradient mapping over Λ and thus the equilibrium problem cannot be reformulated as a standard convex optimization problem.

With the various penalty costs introduced, the cost on arc $a=(i,j)\in A$ incurred by each unit of flow x_a^{gd} is

$$c_a^{gd}(\mathbf{x}) = l(i,j) + \begin{cases} U_{oj}^{odg}, & \text{if } i = o \in O, \\ U_{id}^{g}, & \text{if } j = d \in D, \\ W_a(\mathbf{y}), & \text{if } a \text{ is a boarding or a transfer arc,} \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

The variational problem (18) may be rewritten in terms of the partial arc flows $\mathbf{x}=\{x_a^{gd}\}$ as

$$\text{Find } \mathbf{x}^* \in \Gamma \text{ satisfying } (\mathbf{x}^*-\mathbf{x})^T c(\mathbf{x}^*) \le 0, \quad \forall \mathbf{x} \in \Gamma, \tag{20}$$

where $c(\mathbf{x}) = \{c_a^{gd}(\mathbf{x})\}$ and Γ is the set of feasible flows $\mathbf{x}$.

Most existing solution algorithms for problem (20) require some sufficient conditions on the monotonicity of the mapping $c(\mathbf{x})$ to converge, although these properties seldom hold in practical applications (Fisk and Nguyen 1982). Hence, existing algorithms for the asymmetric traffic equilibrium problem cannot be applied directly here. Nevertheless, good heuristics based on existing methods can be definitely designed to solve problem (20). One possible solution approach is suggested in the next subsection.

3.2. A Solution Method

Since we have a polyhedral ground set Γ, it seems advantageous to apply a simplicial decomposition strategy which has long been employed in mathematical programming (Von Hohenbalken 1977) and in many traffic assignment applications (Hearn et al. 1987, Larsson and Patriksson 1992, Patriksson 1994). This allows us to formulate the nonmonotone variational inequality problem (20) as an equivalent nonlinear minimization problem defined on the convex hull of a subset of extreme points (Smith 1983a, 1983b, Hearn et al. 1984). Let S denote the convex hull of a subset $\{\mathbf{v}^0, \mathbf{v}^1, \dots, \mathbf{v}^k\}$ of extreme points of Γ; consider the variational problem

$$\text{Find } \mathbf{x}^* \in S \text{ satisfying } (\mathbf{x}^*-\mathbf{x})^T c(\mathbf{x}^*) \le 0, \ \forall \mathbf{x} \in S, \tag{21}$$

and the family of Smith's (1983a) functions:

$$Z(\mathbf{x}) = \sum_{i=0}^{k} ([(\mathbf{x}-\mathbf{v}^i)^T c(\mathbf{x})]^+)^\beta, \ \forall \mathbf{x} \in S, \tag{22}$$

for any integer $\beta \ge 1$. It can be shown that

$$Z(\mathbf{x}^*) = \min_{x\in S} Z(\mathbf{x}) = 0.$$

Note that for $2 \le \beta < \infty$, $Z(\mathbf{x})$ is differentiable everywhere and its gradient is

$$\nabla Z(\mathbf{x}) = \sum_{i=0}^{k} \beta([(\mathbf{x}-\mathbf{v}^i)^T c(\mathbf{x})]^+)^{\beta-1}((\mathbf{x}-\mathbf{v}^i)^T \nabla c(\mathbf{x}) + c(\mathbf{x})).$$

In view of the above, consider the following general scheme.

Procedure. *Equilibrium Assignment* (ε, β)

Step 0. (initialization)

Select $\mathbf{v}^0$, set $\Pi := \{\mathbf{v}^0\}$; $\mathbf{x}^0 := \mathbf{v}^0$; $k := 0$.

Step 1. (extreme point/column generation)

Determine

$$\mathbf{v}^{k+1} = \arg\min_{\mathbf{x}\in\Gamma} (\mathbf{x}-\mathbf{x}^k)^T c(\mathbf{x}^k).$$

If $(\mathbf{x}^k - \mathbf{v}^{k+1})^T c(\mathbf{x}^k) \le \varepsilon[(\mathbf{v}^{k+1})^T c(\mathbf{x}^k)]$, stop with $\mathbf{x}^* = \mathbf{x}^k$ (ε-equilibrium).

Otherwise set $\Pi := \Pi \cup \{\mathbf{v}^{k+1}\}$; $k := k+1$.

Step 2. (master problem)

Determine the global minimizer $\boldsymbol{\lambda}^*$ of

$$\min Z(\boldsymbol{\lambda}) = \sum_{i=0}^{k} ([(\mathbf{V}\boldsymbol{\lambda} - \mathbf{v}^i)^T c(\mathbf{V}\boldsymbol{\lambda})]^+)^\beta$$

subject to

$$\sum_{i=0}^{k} \lambda_i = 1$$

$$\lambda_i \ge 0, \ \forall i = 0, 1, \dots, k,$$

where $\mathbf{V} = [\mathbf{v}^0, \mathbf{v}^1, \dots, \mathbf{v}^k]$.

Set $\mathbf{x}^k = \mathbf{V}\boldsymbol{\lambda}^*$. Updated the set Π and go to Step 1.

Note that this scheme may also be interpreted as a column generation scheme. It is well known that the linear programming subproblem of Step 1 reduces to a shortest path problem, for which various very efficient algorithms may be implemented for large scale networks (Gallo and Pallottino 1986, Mondou et al. 1991). It should be noted here that computing shortest paths (with possible static cycles) in a dynamic network is completely equivalent to computing classical shortest paths in the corresponding space-time network (Pallottino and Scutellà 1998). The main difficulty resides rather in the determination of a global minimizer of the generally nonconvex subproblem in Step 2 and the theoretical convergence of the whole procedure depends naturally on the convergence of this step and on the updating rule adopted for the set of extreme points Π. For added insights on these topics we refer the reader to Hearn et al. (1985, 1987).

At first view, it seems that one has to work in the space of partial flows $\{x_a^{gd}\}$ and this will severely limit the size of the problem that can be solved. We will show below that even if the variational problem (20) cannot be written in the space of total arc flows, all the computations needed for the above algorithm can be carried out in that space.

First, note that the definition (19) of arc cost $c_a^{gd}(\mathbf{x})$ allows us to decompose the product $\mathbf{x}^T c(\mathbf{x})$ into four terms,

$$\mathbf{x}^T c(\mathbf{x}) = \sum_{a\in A} l_a y_a + \sum_{a\in A_b} W_a(\mathbf{y}) y_a + \sum_g \sum_{d\in D} \left[\sum_{o\in O} \sum_{(o,j)\in S^+(o)} U_{oj}^{odg} x_{oj}^{gd} \right] + \sum_g \sum_{d\in D} \left[\sum_{(j,d)\in S^-(d)} U_{jd}^{g} x_{jd}^{gd} \right],$$

where $A_b \subset A$ is the set of boarding and transfer arcs. The third term is the total penalty cost due to early departures from all origins. The last term is the total penalty cost due to early or late arrivals at destinations. Note that only the second term contains flow dependent cost functions. This leads to further decomposition. Indeed, let $\{\mathbf{y}^0, \mathbf{y}^1, \ldots, \mathbf{y}^k\}$ denote the total arc flows corresponding to the extreme flows $\{\mathbf{v}^0, \ldots, \mathbf{v}^k\}$. The three other terms associated with a given extreme point $\mathbf{v}^i$ may be collected in a single term Ψ^i:

$$\Psi^i = \sum_{a\in A} l_a y_a^i + \sum_g \sum_{d\in D} \left[\sum_{o\in O} \sum_{(o,j)\in S^+(o)} U_{oj}^{odg} (v_{oj}^{gd})^i \right] + \sum_g \sum_{d\in D} \left[\sum_{(j,d)\in S^-(d)} U_{jd}^{g} (v_{jd}^{gd})^i \right]. \quad (23)$$

It is important to note that Ψ^i, $i = 0, 1, \ldots, k$, can be computed sequentially at the same time as the network loading to produce $\mathbf{v}^i$. This also implies that shortest paths calculations must be carried out from the destination in contrast to the usual origin-based method.

Considering the kth iteration of the above procedure, the computation of the objective function of the subproblem in Step 2 and its gradient can be carried out as follows. For a given $\lambda = \{\lambda_j\}$, let $\bar{\mathbf{y}} = \sum_{j=0}^k \lambda_j \mathbf{y}^j$. Then taking into account the definition (23),

$$Z(\boldsymbol{\lambda}) = \sum_{i=0}^k ([T_i]^+)^\beta = \sum_{i=0}^k ([(\mathbf{V}\boldsymbol{\lambda} - \mathbf{v}^i)^T c(\mathbf{V}\boldsymbol{\lambda})]^+)^\beta$$

can be rewritten as

$$T_i = \sum_{a\in A_b} W_a(\bar{\mathbf{y}})(\bar{y}_a - y_a^i) + \left(\sum_{j=0}^k \Psi^j \lambda_j \right) - \Psi^i. \quad (24)$$

Similarly, the gradient of $Z(\boldsymbol{\lambda})$ may be expressed as

$$\nabla Z(\boldsymbol{\lambda}) = \sum_{i=0}^k \beta([T_i]^+)^{\beta-1} \nabla T_i,$$

where the sth component of ∇T_i can be expressed as

$$(\nabla T_i)_s = \sum_{a\in A_b} \left((\bar{y}_a - y_a^i) \frac{\partial W_a(\bar{\mathbf{y}})}{\partial \lambda_s} + W_a(\bar{\mathbf{y}}) y_a^s \right) + \Psi^s.$$

Recall that for a boarding or transfer arc $a = (i_m, j)$, the penalty cost $W_a(\bar{\mathbf{y}})$ and its gradient $\partial W_a(\bar{\mathbf{y}})/\partial \lambda_s$ can be expressed as

$$W_a(\bar{\mathbf{y}}) = \alpha_a \left(\left[\sum_{\ell=0}^m \bar{y}_{i_\ell j} - \rho_a u_a \right]^+ \right)^\theta,$$

$$\frac{\partial W_a(\bar{y})}{\partial \lambda_s} = \alpha_a \theta \left(\left[\sum_{\ell=0}^m \bar{y}_{i_\ell j} - \rho_a u_a \right]^+ \right)^{\theta-1} \sum_{\ell=0}^m y_{i_\ell j}^s.$$

These developments clearly show that the master problem in Step 2 can be solved in the space of

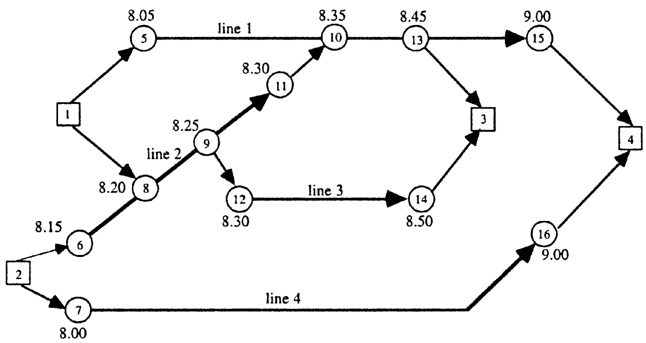

Figure 4 An Illustrative Example

total flows $\{\mathbf{y}^0, \mathbf{y}^1, \ldots, \mathbf{y}^k\}$ and hence only these flows must be stored during the execution of the whole procedure. This is a clear advantage for practical applications.

3.3. An Illustrative Numerical Example

Consider the network in Figure 4 where the access and boarding arcs are combined into a single arc. There are 4 origin-destination pairs $\{(1,3),(1,4),(2,3),(2,4)\}$, and one group of passengers with the arrival time interval $[8h45, 9h00]$. The traffic demand is $\{10, 10, 10, 10\}$, and the capacity of every carrier is $u = 20$. The arc costs are given in Table 1 for $\alpha_a = 1$, $\rho_a = 0.80$ (thus $\rho_a u_a = 16$) for every boarding arc, and $\theta = 2$.

For this application, the subproblem in step 2 with $\beta = 2$ is solved with a linear approximation algorithm and PARTAN direction (see, for example, Luenberger 1984, or Leblanc et al. 1985, Florian et al. 1987, Arezki and Vliet 1990 in relation to the symmetric cost traffic assignment problem). Solving the above assignment problem produces the ε-equilibrium ($\varepsilon = 0.0001$) solution described in Table 2. Note that the flows $y_{8,9} = 21$ and $y_{10,13} = 21.472$ exceed the capacity of the respective arcs ($u = 20$). One may indirectly enforce the capacity constraints by choosing appropriate values for the parameters of the arc penalty functions $(\alpha_a, \rho_a, \theta)$.

It is interesting to note that, for this example, an exact equilibrium solution can be computed as the limiting solution to the above variational problem (21). Furthermore, this equilibrium solution is not a system optimum solution (Table 3) that one would expect to obtain with a standard approach for an equilibrium assignment problem with constant arc costs and arc capacities (Hearn 1980, Patriksson 1994).

Table 1 Arc Costs

Arc	(i, j)	C_a
1	(1, 5)	$15+([x_{1,5}-16]^+)^2$
2	(1, 8)	$([x_{1,8}+x_{6,8}-16]^+)^2$
3	(2, 6)	$([x_{2,6}-16]^+)^2$
4	(2, 7)	$15+([x_{2,7}-16]^+)^2$
5	(5, 10)	30
6	(6, 8)	5
7	(7, 16)	60
8	(8, 9)	5
9	(9, 11)	5
10	(9, 12)	$5+([x_{9,12}-16]^+)^2$
11	(10, 13)	10
12	(11, 10)	$5+([x_{11,10}+x_{5,10}-16]^+)^2$
13	(12, 14)	20
14	(13, 3)	0
15	(13, 15)	15
16	(14, 3)	0
17	(15, 4)	0
18	(16, 4)	0

Table 2 An ε-Equilibrium Solution

Path Flow h_p	Path Description	C_p	l_p
5.236	{1, 5, 10, 13, 3}	55.000	55
0.000	{1, 8, 9, 11, 10, 13, 3}	79.944	25
4.764	{1, 8, 9, 12, 14, 3}	55.000	30
10.000	{1, 5, 10, 13, 15, 4}	70.000	70
0.000	{1, 8, 9, 11, 10, 13, 15, 4}	94.944	40
0.000	{2, 6, 8, 9, 11, 10, 13, 3}	60.000	30
10.000	{2, 6, 8, 9, 12, 14, 3}	35.056	35
6.236	{2, 6, 8, 9, 11, 10, 13, 15, 4}	75.000	45
3.764	{2, 7, 16, 4}	75.000	75

Table 3 Equilibrium and System Optimum Solution

Equil. h_p	Sys. Opt. h_p	Path Description	C_p	l_p
5	0	{1, 5, 10, 13, 3}	55	55
0	10	{1, 8, 9, 11, 10, 13, 3}	80	25
5	0	{1, 8, 9, 12, 14, 3}	55	30
10	10	{1, 5, 10, 13, 15, 4}	70	70
0	0	{1, 8, 9, 11, 10, 13, 15, 4}	95	40
0	0	{2, 6, 8, 9, 11, 10, 13, 3}	60	30
10	10	{2, 6, 8, 9, 12, 14, 3}	35	35
5	0	{2, 6, 8, 9, 11, 10, 13, 15, 4}	75	45
5	10	{2, 7, 16, 4}	75	75

3.4. Conclusion

We have developed in this paper a graph-theoretic framework for the passenger assignment problem that encompasses simultaneously the departure time and the route choice dimensions. The proposed dynamic framework is built upon the new concept of *path available capacity* that allows us to capture the flow priority induced by the implicit FIFO rule that binds passengers at every access point to the transit network. No such general framework has been developed to date, and existing frameworks for urban networks are not directly applicable since they are intrinsically static.

From the dynamic framework, distinctive features of the passenger routing are investigated and illustrated. A passenger equilibrium flow model is then defined and a mathematical formulation suggested. A computational algorithm for the determination of an equilibrium flow based on a penalty approach is also described. Development and analysis of convergent algorithms for large scale networks based on the described general procedure are in progress, and an integration of the present model with that in Nguyen and Pallottino (1988) seems to be a natural extension.

References

Alfa, A. S. 1986. A review of models for the temporal distribution of peak traffic demand. *Transportation Res.* **20B**(6) 491–499.

——. 1989. Departure rate and route assignment of commuter traffic during peak period. *Transportation Res.* **23B**(5) 337–344.

Arezki Y., D. Van Vliet. 1990. A full analytical implementation of the PARTAN/Frank-Wolfe algorithm for equilibrium assignment. *Transportation Sci.* **24**(1) 58–62.

Carey, M. 1992. Nonconvexity of the dynamic traffic assignment problem. *Transportation Res.* **26B**(2) 127–133.

——, A. Srinivasan. 1988. Congested network flows: Time-varying demands and start-time policies. *Eur. J. Oper. Res.* **36**(2) 227–240.

Dafermos, S. 1980. Traffic equilibria and variational inequalities. *Transportation Sci.* **14**(1) 42–54.

De Palma, A., M. Ben-Akiva, C. Lefevre, N. Litinas. 1983. Stochastic equilibrium model of peak period traffic congestion. *Transportation Sci.* **17**(3) 430–453.

Fisk, C. S., S. Nguyen. 1982. Solution algorithm for network equilibrium models with asymmetric user costs. *Transportation Sci.* **16**(3) 361–381.

Florian, M., J. Guélat, H. Spiess. 1987. An efficient implementation of the PARTAN variant of the linear approximation method for the network equilibrium problem. *Networks* **17**(3) 319–339.

Friesz, T. L., D. Bernstein, T. E. Smith, R. L. Tobin, B. W. Wie. 1993. A variational inequality formulation of the dynamic network user equilibrium problem. *Oper. Res.* **41**(1) 179–191.

——., F. J. Luque, R. L. Tobin, B. W. Wie. 1989. Dynamic network traffic assignment considered as a continuous time optimal control problem. *Oper. Res.* **37**(6) 893–901.

Gallo G., S. Pallottino. 1986. Shortest path methods: A unifying approach. *Math. Programming Stud.* **26** 38–64.

Gendreau, M. 1984. Une étude approfondie d'un modèle d'équilibre pour l'affectation des passagers dans les réseaux de transport en commun, Ph.D. thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal, Montreal, Canada.

Hearn, D. W. 1980. Bounding flows in traffic assignment models. Research Report 80-4, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL.

——, S. Lawphongpanich, S. Nguyen. 1984. Convex programming formulations of the asymmetric traffic assignment problem. *Transportation Res.* **18B**(4) 357–365.

——, ——, J. A. Ventura. 1985. Finiteness in restricted simplicial decomposition. *Oper. Res. Lett.* **4** 125–130.

——,——,——. 1987. Restricted simplicial decomposition: computation and extensions. *Math. Programming Stud.* **31** 99–118.

Hendrickson C., E. Plank. 1984. The flexibility of departure times for work trips. *Transportation Res.* **18A**(1) 25–36.

Kinderlehrer D., G. Stampacchia. 1980. *An Introduction to Variational Inequalities and Applications*. Academic Press, New York.

Larsson T., M. Patriksson. 1992. Simplicial decomposition with disaggregated representation for the traffic assignment problem. *Transportation Sci.* **26**(1) 4–17.

Leblanc, L. J., R. V. Helgason, D. E. Boyce. 1985. Improved efficiency of the Frank-Wolfe algorithm. *Transportation Sci.* **19**(4) 445–462.

Luenberger, D. G. 1984. *Introduction to Linear and Nonlinear Programming.* Addison-Wesley, Reading, MA.

Mahmassani, H. S., G. L. Chang. 1986. Dynamic aspect of departure-time choice behaviour in a commuting system: Theoretical framework and experimental analysis. *Transportation Res. Record* **1037** 88–101.

——, ——. 1987. On boundedly rational user equilibrium in transportation systems. *Transportation Sci.* **21**(2) 89–99.

——, R. Herman. 1984. Dynamic user equilibrium departure time and route choice on idealized traffic arterials. *Transportation Sci.* **18**(4) 362–384.

Marcotte, P., S. Nguyen. 1998. Hyperpath formulations of traffic assignment problems. P. Marcotte, S. Nguyen, eds. *Equilibrium and Advanced Transportation Modelling.* Kluwer Academic Publisher, Assinipi Park, Norwell, MA, 175–199.

Mondou, J. F., G. T. Crainic, S. Nguyen. 1991. Shortest path algorithms: A computational study with the C programming language. *Comput. Oper. Res.* **18**(8) 767–786.

Newell, G. F. 1987. The morning commute for nonidentical travellers. *Transportation Sci.* **21**(2) 74–88.

Nguyen, S., S. Pallottino. 1988. Equilibrium traffic assignment for large scale transit networks. *Eur. J. Oper. Res.* **37**(2) 176–186.

Pallottino, S., M. G. Scutellà. 1998. Shortest path algorithms in transportation models: classical and innovative aspects. P. Marcotte, S. Nguyen, eds. *Equilibrium and Advanced Transportation Modelling*. Kluwer Academic Publisher, Assinipi Park, Norwell, MA, 245–281.

Patriksson, M. 1994. *The Traffic Assignment Problem: Models and Methods*. VSP, Utrecht, The Netherlands.

Smith, M. J. 1979. The existence, uniqueness and stability of traffic equilibria. *Transportation Res.* **13B**(4) 295–304.

——. 1983a. An algorithm for solving asymmertic equilibrium problems with a continuous cost-flow function. *Transportation Res.* **17B**(5) 365–371.

——. 1983b. The existence and calculation of traffic equilibria. *Transportation Res.* **17B**(4) 291–303.

Spiess, H., M. Florian. 1989. Optimal strategies: A new assignment model for transit networks. *Transportation Res.* **23B**(2) 83–102.

Sumi, T., Y. Matsumoto, Y. Miyaki. 1990. Departure time and route choice of commuters on mass transit systems. *Transportation Res.* **24B**(4) 247–262.

Von Hohenbalken, B. 1977. Simplicial decomposition in nonlinear programming algorithms. *Math. Programming* **13** 49–68.

Wardrop, J. G. 1952. Some theoretical aspects of road traffic research. *Proc. of the Institute of Civil Engineers*, Part 2, No. 1, 325–378.

Wu, J. H., M. Florian, P. Marcotte. 1994. Transit equilibrium assignment: A model and solution algorithms. *Transportation Sci.* **28**(3) 193–203.

Received: April 1995; revisions received: November 1998; accepted: February 2000.