Schedule-Based Transit Assignment Model with Travel Strategies and Capacity Constraints

Younes Hamdouch* Siriphong Lawphongpanich†

*Department of Business Administration
United Arab Emirates University
Al Ain, United Arab Emirates
Younes.Hamdouch@uaeu.ac.ae

†Department of Industrial and Systems Engineering
University of Florida
Gainesville, Florida 32611 USA
Lawphong@ise.ufl.edu

1 Introduction

In this paper, we propose a schedule-based transit assignment model that uses a monolithic time-expanded network to describe passenger and vehicle movements over time. The model assumes that passengers use travel strategies. The latter can be adaptive over time and reduce to simple paths in systems with infrequent services. When loading passengers either randomly or on a First-Come-First-Serve (FCFS) basis, the model takes into account vehicle capacities explicitly and individually. After loading, the percentages of passengers on-board different transit lines and of those who have to wait can be interpreted as diversion and failure-to-board probabilities, respectively, in some frequency-based models (see, e.g., [2], and [9]).

To find a user equilibrium assignment, we formulate the problem as a variational inequality involving a vector-valued function of expected strategy costs. We show that a solution to the variational inequality exists and propose a method based on successive averages to find it. In [4], we demonstrate empirically that the method converges to a user equilibrium solution and discuss its characteristic.

For the remainder, Section 2 discusses the transformation of a static route network into a time-expanded version. Section 3 describes how to specify valid travel strategies. Section 4 shows how to calculate the cost associated with a strategy using the node and arc access probabilities defined in, e.g., [5] and [6]. Section 5 formulates the transit assignment problem as a variational inequality and proposes an algorithm that is based on successive averages and generates strategies as it progresses.

2 Passenger Movements

This section describes two networks. One is the route network that mainly depicts the routes of all transit lines in a static manner and the other is its time-expanded version that displays all possible routes passengers can travel to their destinations at various times.

2.1 Route Network

In a route network, nodes are of two types, one consists of origins and destinations and the other corresponds to transit stations where a transit vehicle stops to load and unload passengers. For emphasis, we also refer to nodes representing transit stations as *station nodes* or, more simply, *stations.*

We assume that a passenger walks from his or her origin to a transit station, use the transit system to arrive at another transit station, and walks from there to his or her destination. With this assumption, arcs in a route network are also of two types. One corresponds to walking and the other represents route segments of transit lines between two consecutive stations. To illustrate, Figure 1 displays a system with an origin node q, a destination node r, and three transit lines. Nodes labeled a, b, c and d are station nodes. In this example, there are two walk arcs, an access arc (q, a) and an egress arc (d, r). The remaining arcs correspond to route segments of the three transit lines.

Parameters associated with each arc (i, j) in the route network consist of a transit fare (v_{ij}), a travel time (c_{ij}) and a capacity (u_{ij}). In Figure 1, both walk arcs, (q, a) and (d, r) have zero transit fare and a unit travel time. We assume that their capacities are infinite. The remaining arcs are segments of transit lines and their parameters are as shown in the figure. For each transit line, the capacities for its route segments are the same and equal to the capacities of the transit vehicles serving the line. Because of its static nature, the route network cannot distinguish, e.g., passengers leaving their origins and transit vehicles starting their routes at different times.

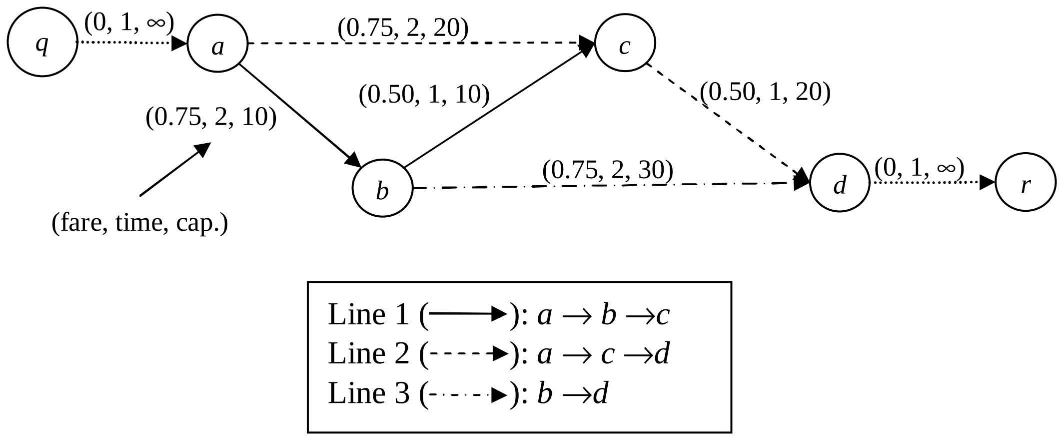

Figure 1: A route network with three transit lines

Line 1			Line 2			Line 3	
a	b	c	a	c	d	b	d
1	3	4	1	3	4	2	4
3	5	6	4	6	7	5	7
5	7	8	7	9	10	8	10
7	9	10					

Table 1: Station departure times for the transit lines in Figure 1

2.2 Time-Expanded Network

Associated with each transit line, there is a schedule listing the times at which a transit vehicle must leave its starting station as well as the scheduled arrival or departure times at stations along its route throughout each day. One method for incorporating such temporal information is by using the time-expanded network (see, e.g., [1] and [6]).

Let $[0, T]$ denote the operating interval of the transit system, i.e., 0 and T represent the start and end of the daily operating hours. In a time-expanded (TE) network, the interval $[0, T]$ is represented as a set of discrete points of the form $\Gamma = \{0, \delta, 2\delta, 3\delta, \cdots, n\delta\}$, where $\delta = T/n$ and n is a positive integer. To simplify our presentation, we ignore δ and simply write Γ as $\{0, 1, 2, 3, \cdots, n\}$. We also assume that all times (arrivals, departures, travel times, etc.) are in multiples of δ.

A node in a TE network (or a TE node) has two labels. One label represents a node in the route network. The other label is an element from the set Γ. In general, each node i in the route network is expanded into $(n + 1)$ nodes of the form i_t, where $t = 0, 1, \ldots, n$, in the TE network. As in the route network, movement arcs in a TE network are of two types, in-vehicle and walk arcs. When a transit vehicle arrives at node i at time t and traverses arc (i, j) in the route network, this movement corresponds to the in-vehicle arc $(i_t, j_{(t+c_{ij})})$ where, as defined in the previous section, c_{ij} denotes the travel time for arc (i, j) in the route network. In general, a route segment (i, j) in the route network is expanded into $(\lfloor T/f_{ij} \rfloor + 1)$ TE arcs of the form $(i_t, j_{(t+c_{ij})})$, where $t = 0, 1, \ldots, \lfloor T/f_{ij} \rfloor$, and f_{ij} is the frequency of the transit line serving the segment. Walk arcs (q, r) and (i, r) in the route network are similarly expanded into $(q_t, j_{(t+c_{qj})})$ and $(i_t, r_{(t+c_{ir})})$, respectively, where i and j are station nodes and $t = 0, 1, \ldots, n$. As before, we also refer to $(q_t, j_{(t+c_{qj})})$ and $(i_t, r_{(t+c_{ir})})$ as *access* and *egress* arcs. In addition to movement arcs, there are wait arcs of the form (i_t, i_{t+1}) that represent passengers having to wait at station i from time t to $(t + 1)$.

To illustrate, Table 1 lists the departure times for each transit line in Figure 1. Using to the schedule in Table 1, Figure 2 displays all possible routes between q and r with a start time between 0 and 3.

There is no explicit travel time associated with each arc in the TE network. The travel time for each TE arc can be deduced from the time indices. For example, the travel time associated with arc (a_1, c_3) is 3 - 1 = 2. On the other hand, all arcs in the TE network have associated transit fares, penalties (measured in monetary units) for, e.g., leaving too early and arriving too late, and capacities. First, all in-vehicle arcs have zero penalties. The transit fares and capacities associated with these arcs are the same as those in the route network. Second, all wait arcs have zero fares, zero penalties and infinite capacities. Finally, all access and egress arcs have zero fares and infinite capacities. However, the penalties of many access and egress arcs are

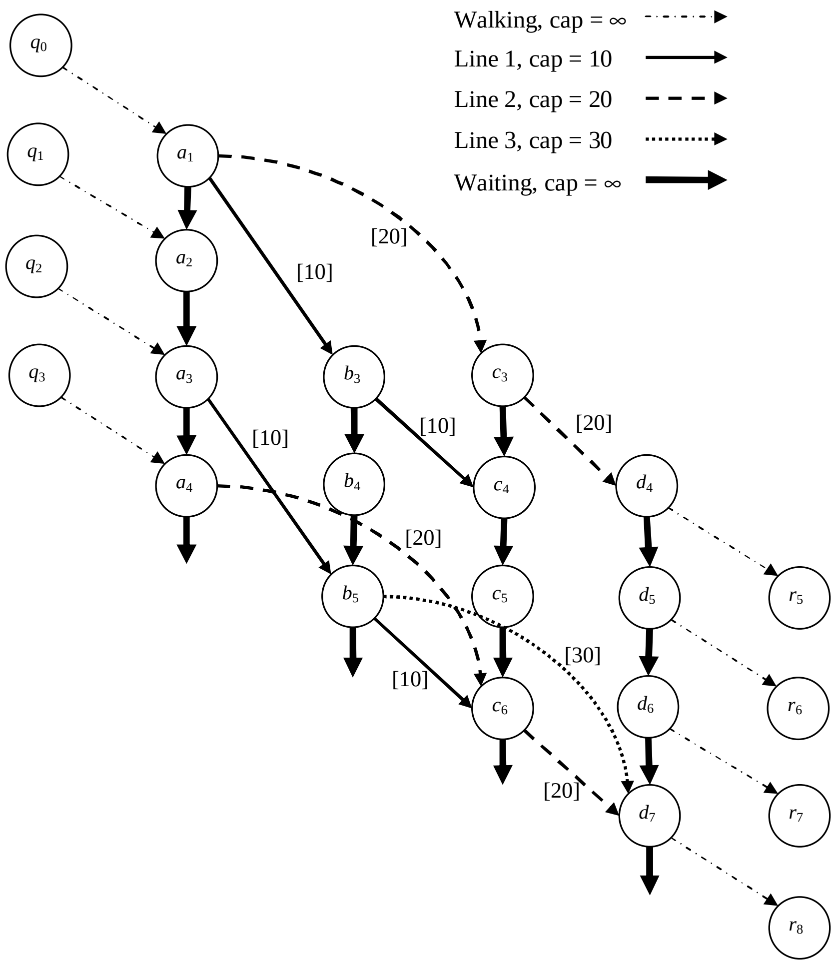

Figure 2: The time-expanded network of the route network in Figure 1

positive and generally different for different passengers, e.g., because penalties depend on each passenger's desired arrival and departure times.

3 Travel Strategies and Generalized Hyperpaths

For each origin-destination (OD) pair (q, r), passengers are divided into groups where each is distinguished by the desired arrival time interval of the passengers in the group. For passengers in group g, the number of passengers in and the desired arrival time interval for the group are denoted as $d^g_{(q,r)}$ and $[t^-_{(q,r)}(g), t^+_{(q,r)}(g)]$, respectively. We assume that passengers use strategies when traveling. To specify a strategy, passengers provide at each node in the TE network a *successor set* or set of immediate successor nodes that enable them to reach their destinations. The order of nodes in each successor set indicates the passenger's preference in their choice of transit lines or their decision to walk in order to advance toward their destinations. Typically, there is only one successor set for each TE node. However, it is possible to have several successor sets at a given node, where each set indicates a different preference when a passenger arrives at node i_t at a different time $\tau \leq t$. Nodes irrelevant to the strategy or OD pair have no successor set. For OD pair (q, r), a strategy for a journey starting at time $\tau \in \Gamma$ is *valid* if the following hold:

i) For all t, there is no successor set at node r_t.

ii) The successor sets at node q_τ must be a subset of the forward-star set, $I^+(q_\tau)$. For all $t \in \Gamma$ and $t \neq \tau$, the successor set at q_t is empty.

iii) The successor sets at every node must induce an acyclic subnetwork of the TE network such that node q_τ has no predecessor and, for every node i_t with nonempty successor sets and $i \notin \{q, r\}$, there is a directed path from q_τ to r_θ, for some $\theta \in \Gamma$ and $\theta > \tau$, that passes through node i_t.

The induced subnetwork in condition (iii) is a generalization of the hyperpath representation of travel strategies (see, [7] and [11]). We refer to the subnetwork herein as a "generalized hyperpath".

4 Expected Strategy Cost

To calculate the expected cost for each strategy, let $S_{(q,r)}$ denote the set of all strategies for OD pair (q, r). Thus, an element of $S_{(q,r)}$ is of the form $s^{\tau}_{(q,r)}(m)$ denoting the m-th strategy for OD pair (q, r) that starts at time τ. However, it is more convenient to also refer to elements of $S_{(q,r)}$ as s instead of $s^{\tau}_{(q,r)}(m)$. Using this simplified notation, let $x^s_{(q,r,g)}$ be the number of passengers for OD pair (q, r) in group g using or assigned to strategy s and X be a vector with $x^s_{(q,r,g)}$ as its components, i.e., X is a strategy assignment (SA) vector. Mathematically, the set of all feasible SA vectors can be stated as follows:

$$\mathcal{X} = \{X : \textstyle\sum_{s \in S_{(q,r)}} x^s_{(q,r,g)} = d^g_{(q,r)}, \forall (q, r, g)\}.$$

For a given $X \in \mathcal{X}$, $\pi^{s,\tau}_{(i_t, j_{(t+c_{ij})})}(X)$ is the (conditional) probability that a passenger using strategy s will traverse arc $(i_t, j_{(t+c_{ij})})$ given that he or she arrives node i at time $\tau \leq t$. When $(i_t, j_{(t+c_{ij})})$ corresponds to a segment of a transit line, $\pi^{s,\tau}_{(i_t, j_{(t+c_{ij})})}(X)$ is the probability that a passenger (or, intuitively, a proportion of passengers) using strategy s can board the transit vehicle serving the arc, where boarding can be accomplished in either FCFS or random manner. Conceptually, these access probabilities are similar to the "diversion probabilities" in, e.g., [9]. On the other hand, $\pi^{s,\tau}_{(i_t, i_{(t+1)})}(X)$ is similar to the "failure-to-board" probability in, e.g., [2], and corresponds to the probability that a passenger using strategy s arriving at node i at time $\tau \leq t$ has to wait at a station i.

Similarly, $\beta^{s,\tau}_{it}(X)$ is the probability that a passenger using strategy s and arriving at station i at time $\tau \leq t$ has to remain at the node at least until time t. For convenience, we also refer to $\pi^{s,\tau}_{(i_t, j_{(t+c_{ij})})}(X)$ and $\beta^{s,\tau}_{it}(X)$ as the arc and node access probabilities, respectively. Both depend on the strategy assignment vector X because, e.g., different strategy assignments may led to different transit vehicles being full when arriving at station node i at time t. This would prevent some passengers from boarding, thereby forcing them to use a wait arc and delaying them from reaching their destinations. The procedure for determining these probabilities involves *loading* the TE network according to a given strategy assignment vector X and is an extension to the ones in [5] and [6]. See [4] for details.

Given the node and arc access probabilities, the expected strategy cost is given by the following expression:

$$\begin{aligned} C^s_{(q,r,g)}(X) = \; & \sum_{t=0}^{n} \sum_{\tau=0}^{t} \sum_{(i,j)\in A} (\gamma c_{ij} + v_{ij} + p^g_{(i_t, j_{(t+c_{ij})})}) \beta^{s,\tau}_{i_t}(X) \cdot \pi^{s,\tau}_{(i_t, j_{(t+c_{ij})})}(X) \\ & + \sum_{t=1}^{n} \sum_{\tau=1}^{(t-1)} \gamma \beta^{s,\tau}_{i_t}(X) \cdot \pi^{s,\tau}_{(i_t, t_{(t+1)})}(X) \end{aligned} .$$

where c_{ij}, v_{ij}, and $p^g_{(i_t, j_{(t+c_{ij})})}$ are the travel time, transit fare, and penalty for $(i_t, j_{(t+c_{ij})})$, respectively, and γ is a factor converting time into monetary units. (See [4] for details concerning the penalty for $(i_t, j_{(t+c_{ij})})$). The second summation is the expected cost associated with waiting. In the above expression, node and arc probabilities associated with events occur prior to the start time of a strategy s are irrelevant and assumed to be zero.

5 User Equilibrium

We say that a SA vector X^* is in a user equilibrium if no passenger has any incentive to change his or her strategy based on expected strategy costs. Using an argument similar to [10] and [3], X^* is in a user equilibrium if and only if X^* solves the following variational inequality (denoted as VI$[C(X), \mathcal{X}]$):

$$C(X^*)^T (X - X^*) \geq 0, \quad \forall X \in \mathcal{X},$$

where $C(X)$ is a vector of expected strategy costs associated with X. Then, the following results follow immediately from those in [5].

Theorem 1 *$C(X)$ is lower semi-continuous on $\mathcal{X}$.*

Theorem 2 *VI$[C(X), \mathcal{X}]$ has at least one solution.*

Instead of enumerating all possible travel strategies a priori, below is a method that generates them by solving a dynamic program during each iteration and use successive averages of previously generated strategies as the current solution to VI$[C(X), \mathcal{X}]$.

Method of Successive Averages

Step 0: For each triplet (q, r, g), select an initial strategy $s[1]$ and set $x^{s[1]}_{(q,r,g)} = d^g_{(q,r)}$ and $x^s_{(q,r,g)} = 0, \forall s \neq s[1]$. Let $X^{[\alpha]}$ be the associated SA vector. Set $\alpha = 1$ and go to Step 1.

Step 1: For each triplet (q, r, g), let $s[\alpha]$ be a least (expected) cost strategy with respect to $X^{[\alpha]}$ and set $y^{s[\alpha]}_{(q,r,g)} = d^g_{(q,r)}$ and $y^s_{(q,r,g)} = 0, \forall s \neq s[\alpha]$. Let $Y^{[\alpha]}$ be the associated SA vector.

Step 2: If $\frac{C(X^{[\alpha]})^T (X^{[\alpha]} - Y^{[\alpha]})}{C(X^{[\alpha]})^T X^{[\alpha]}} \leq \varepsilon$, stop. Otherwise, set $X^{[\alpha+1]} = \frac{1}{(\alpha+1)} (\alpha X^{[\alpha]} + Y^{[\alpha]})$ and $\alpha = \alpha + 1$. Go to Step 1.

For each combination (q, r, g), strategy $s[1]$ in Step 0 can correspond to leaving node q at time $L^g_{(q,r)}$, the latest departure time (see, e.g., [8]), and taking a shortest path to r in the TE network. In Step 2, each component of $C(X^{[\alpha]})$, or $C^s_{(q,r,g)}(X^{[\alpha]})$, can be determined using the loading procedure described in [4]. To determine the least cost strategy $s[\alpha]$ in Step 1, we use a dynamic program or Bellman's equation similar to the one in [6]. See [4] for details and an example illustrating the above algorithm.

Acknowledgements

This research was partially supported by grants from the National Science Foundation (DMI-0300316) and Volvo Research and Educational Foundations (SP-2004-5).

References

[1] Ahuja, R.K., Magnanti, T.L., Orlin, J.B., 1993. Network Flows: Theory, Algorithms, and Applications, Prentice Hall, NJ.

[2] Bell, M.G.H., Schmocker, J.-D., 2004a. A solution to the congested transit assignment problem. In Wilson, N.H.M., Nuzzolo, A. (Eds.), Scheduled-Based Dynamic Transit Modeling: Theory and Applications, Springer, New York, 263 - 280.

[3] Dafermos, S., 1980. Traffic equilibrium and variational inequalities. Trans. Sci., 14, 42 - 54.

[4] Hamdouch, Y., Lawphongpanich, S., 2006. Schedule-based transit assignment model with travel strategies and capacity constraints. Technical Report, Dept. of Business Administration, UAE University, Al Ain, UAE.

[5] Hamdouch, Y., Marcotte, P., Nguyen, S., 2004a. Capacitated transit assignment with loading priorities. Math. Prog. B, 101, 205 - 230.

[6] Hamdouch Y., Marcotte, P., Nguyen, S., 2004b, A strategic model for dynamic traffic assignment. Networks and Spatial Econ., 4, 291 - 315.

[7] Nguyen, S., Pallottino, S., 1988. Equilibrium traffic assignment for large scale transit networks. European Journal of Operational Research, 37(2), 176 - 186.

[8] Nguyen, S., Pallottino, S., Malucelli, F., 2001. A Modeling framework for passenger assignment on a transport network with time tables. Trans. Sci., 35(3). 238 - 249.

[9] Nuzzolo, A., 2003. Transit path choice and assignment model approaches. In Lam, W.H.K., Bell, M.G.H. (Eds.), Advanced Modelling for Transit Operations and Service Planning, Pergamon, Oxford, 93 - 124.

[10] Smith, M.J., 1979. Existence, uniqueness, and stability of traffic equilibria. Transportation Research Part B, 13, 295 - 304.

[11] Spiess, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. Transportation Research Part B, 23(2), 83 - 102.