Transit Schedule Design in Dynamic Transit Network with Demand and Supply Uncertainties

Yuqing ZHANG
Research Student
Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hong Kong
Fax: 852-2334-6389
E-mail: 07902133r@polyu.edu.hk

William H. K. LAM
Chair Professor
Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hong Kong
Fax: 852-2334-6389
E-mail: cehklam@polyu.edu.hk

Agachai SUMALEE
Assistant Professor
Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hong Kong
Fax: 852-2334-6389
E-mail: ceasumal@polyu.edu.hk

Abstract: This paper proposed a novel transit schedule design model in dynamic transit network with uncertainties. An array of uneven headways for each line is designed to optimize the integrated transit service efficiency and reliability. It is formulated as a bi-level problem, by which the transit schedule can be varied so as to make response to the current demand. The objective of the upper level model is to optimize the integrated transit network efficiency and reliability. The lower-level model is a reliability-based transit assignment model, in which uncertainties from both demand and supply side are taken into consideration. The bi-level problem is solved by applying the Genetic algorithm (GA). Finally, a numerical example is used to show the performance of the proposed model and solution algorithm.

Keywords: transit schedule design, reliability-based dynamic transit assignment, Genetic Algorithm

1. INTRODUCTION

The growth of city population density and the high mobility requirement calls for efficient and reliable public transit services. The setting of transit lines and frequencies are the two main components of transit network design for the improvement of the level of service. However the two planning techniques serve respective planning purposes. The determination of transit routes or lines aims at long-term planning as the change of transit lines, especially at the network level, is only implemented in long term period. Nevertheless, the setting of frequency or headway serves the short-term planning purpose of flexible adaptation for the day-to-day and within-day passenger demand variation.

The time-varying passenger demand often causes uncertainties in the congested transit network. Besides, uncertainties in transit network may also stems from adverse weather or traffic incidents on road. These uncertainties from both demand and supply side affect transit service severely and cause the vehicle double-heading, knocking-on (bunching) and overtaking phenomenon. The unexpected long waiting time and overload delay for passenger severely impact transit service of reliability. The efficiency is another significant concern of transit agencies. The empty seat in the dispatching vehicle is redundant and will result in low productivity of a run. Thus the decision of transit scheduling should embed not only the
average situation but also the reliability component for a more comprehensive planning.

In this paper, the objective of transit scheduling is to minimize the integrated value of network travel time and uncertainties while the number of vehicles dispatched is minimized. In literature, the setting of a reasonable timetable has been studied not only from saving number of operating vehicles (Ceder, 2003; Gao et al., 2004; Uchida et al., 2007), but also from improving the passenger level of service, such as maximizing social benefit (Furth and Wilson, 1981), maintaining the headway regulation (Ding and Chien, 2001), improving the line connection and timetable synchronization (Ceder et al., 2001; Fleurrent et al., 2005), etc. Among these, the bi-level model has been developed by Gao et al. (2004) and Uchida et al. (2007) consider the service demand interaction and managed to carry out the sensitivity analysis to determine the transit frequencies. The frequency for each line is a fixed number and transit assignment model in the lower level is static, which implies these models considers the demand and service in an average view. However, the flexibility of transit scheduling and the agility of reflecting the variation of demand needs reconstruction of the bi-level transit scheduling model.

The bi-level model proposed in the paper is a reliability-based dynamic transit scheduling model. The lower-level model in this paper is a schedule-based transit assignment model in network with uncertainties, which produce the time-varying passenger demand and the mean and variance of network passenger travel cost. The time-space expended network enables the upper-level scheduling model setting the timetable unevenly. The embedding of reliability component allows transit agencies to directly take service reliability into consideration by implementing this model. However, setting the uneven headway between dispatching vehicles is too complicated for mathematical programming because of the large number of variables and constraints. Ceder (2003) proposed three graphical procedures to determine the balanced (uneven) headway in terms of passenger count and patronage of a certain transit line over a time period. However, the passenger travel decision and transit service attributes interaction is not considered in these procedures.

The mathematical programming or optimization method is applicable when dealing with even headway problem though the convergence is rarely discussed (Gao et al., 2004; Uchida et al., 2007). However, for alternative headways, the heuristic graphical methods (Ceder, 2003; Yan and Chen, 2002) are used instead of analytical methods to find a feasible and reasonable solution. Genetic Algorithm is regarded providing robust search as well as a near optimal solution in a reasonable time. Hence it is widely applied for transit route design and scheduling problems (Pattnaik et al., 1998; Kidawí et al., 2005; Shrivastava and O'Mahon, 2006), and is capable to solve the bi-level problem with large variable dimension. Genetic algorithms (GA) are adaptive heuristic search algorithms which evolve by natural selection and produce the solution. In this paper, GA is applied to find out the headways of each line, which minimize the integrated value of travel cost while minimizing the dispatched number of vehicles.

This paper is organized as follows. In section 2, the reliability-based stochastic user equilibrium (RSUE) model is described for the lower level of the scheduling model. The upper-level of the scheduling model is then proposed in section 3, and solved by genetic algorithm with the intelligence of assigning transit demand and dealing with scheduling constraints in section 4. The numerical example is carried out to show the performance of the model in section 5. The concluding remarks are given in Section 6.
2. LOWER LEVEL MODEL: RSUE

2.1 Network Representation
The transit service network is normally represented by either time-space trajectory graph (Powell and Sheffi, 1983; Ceder, 2007), time-space diachronic model (Nuzzolo et al., 2001), and time-extended model (Hamdouch and Lawphongpanich, 2008). Among these, the time-space trajectory model is capable of illustrating the vehicle dispatching, line run time, the schedule adherence, and the vehicle encounter or overtaking on the run. As shown in Figure 1, the even headways may not be maintained through the line. On the contrary, the uneven headways, which consider the demand variation overtime, are better to maintain headways irregularity, and other service attributes.

Given a transit network $\Omega(I,J,L)$, the $i^{th}$ transit vehicle of line $l$ at $j^{th}$ stop, $V_{i,j,l}$, can uniquely define the related arrival time and departure time: $T_{i,j,l}^a$, $T_{i,j,l}^d$. The $i^{th}$ vehicle and $(i+1)^{th}$ vehicle of the same line meets at the $j^{th}$ stop due to the variations in dwelling time and on-road travel time. The stochastic vehicle on-road time due to traffic accident and incidence on road (Chen et al., 1999) or the adverse weather (Lam et al., 2008) is evident in the real word and has been studied extensively. However, the stochastic vehicle dwelling time is more complicated for modeling.

2.2 Modeling Transit Demand
In this paper, the stochastic passenger demand is assumed to follow the Normal distribution. $D_{rs}$ denotes the passenger demand between an origin-destination (OD) pair $r-s$ during the
modeling time horizon of $[0,T]$. For each time interval $\tau_i \in (\ldots,\tau_{i-1},\tau_i,\tau_{i+1},\ldots)$, we further assume that the number of passengers arrival at each time interval $\tau_i$ for each line $l$ follows inhomogeneous Poisson process $\{Q_l(\tau), \tau \geq 0\}$. According to the central limit theorem, the total number of passengers $D_l$ for each line $l$ over the whole period can be approximated by $Q_l(\tau)$. The passenger demand for all OD and for all lines is then defined as:

$$\sum_l (D_l - A_l) = \sum_{r,s} D^{rs},$$

where $A_l$ is the transfer passenger of each line. A transfer incidence matrix is generated relating to passenger’s route choice decision.

2.3 Modeling Transit Services

On the arrival of the transit vehicle if the vehicle has spare capacity, the passenger waiting in the queue will start boarding. When both the time interval between the arrival of every two successive passengers and the time for passenger boarding are stochastic, the number of passengers getting aboard and the total boarding time for the process can hardly be obtained if the mean of passenger boarding time and arrival rate are not known. Nevertheless, these mean values are not adequate for representing transit network attributes, especially when the risk-taking behavior of passenger is considered in transit network with uncertainties.

To model the boarding process, the M/G/1 queue theory is applied to get the analytical expression of the stochastic boarding time of passengers, thereby the stochastic dwelling time and run headway of transit vehicles. Here we directly refer to the Markovian property of the passenger arrival and boarding (PAB) process in Zhang et al. (2009), and the definitions in M/G/1 queues.

At the arrival of transit vehicle, if there is more than one passenger waiting at the initial time of boarding process, the M/G/1 queue can not be applied directly. To interpret this period by PAB process, the arrival rate of passenger is assumed infinite, i.e., the arrival interval of passenger is approximately zero. Thus the traffic intensity of PAB process $\rho > 1$, and the mean of busy period (time of boarding/ vehicle dwelling time) is infinite, as well as the number of passengers getting aboard. However, the finite system is bundled by capacity constraint of transit vehicle, the number of passengers waiting, and the time bundle of Poisson arrival process with different arrival rates:

$$E[B(\tau_i)] = \tau_i, \quad E[N(\tau_i)] = 1/E[B_{per}] .$$

In reality, the arrival rate of passengers is not constant over a period of time, particularly when the passenger departure time should be varied in practice. But it is sensible to assume a constant passenger arrival rate for a small time interval (Clark and Watling 2006). Within each time interval, consider the each time for boarding as renewal interval, by the Renewal theory, both the mean and variance of passenger boarding time can be derived:

$$E[B(\tau_i)] = \int_0^\infty (x) dG^{N(\tau_i)}(x)$$

$$\text{var}[B(\tau_i)] = \int_0^\infty (x^2) dG^{N(\tau_i)}(x^2) + [\int_0^\infty (x) dG^{N(\tau_i)}(x)]^2$$
Note that the coincidence of the mean of boarding time with Equation 2 in this phase:
\[
E[B(\tau_i)] = \int_0^\infty (x) dG^{N(\tau_i)}(x) = E[B_{\text{per}}^N(\tau_i)] = E[N(\tau_i)] \cdot E[B_{\text{per}}] = \tau_i.
\] (6)

At the beginning of a time interval, there may be more than one passenger in the queue, but the number of passengers queuing in addition to the passenger who will arrive in the interval is less than \( E[N(\tau_i)] \). Define this time interval as Phase 2, and the time periods described above as Phase 1. Simplify the notion of passenger getting aboard at Phase 1 as:
\[
n_i = \sum_i E[N(\tau_i)] = \sum_i \frac{i}{E[B_{\text{per}}]}.
\] (7)

The PAB process will end at Phase 2, and the number of passengers getting aboard is:
\[
n_j = \lambda(\tau_j) + q'_j,
\] (8)
where \( q'_j \) is the number of passengers waiting at the beginning of interval \( \tau_j \). Applying the same logic above, the mean and variance of passenger boarding time at this interval is:
\[
E[B(\tau_j)] = \int_0^\infty (x) dG^{n_j}(x)
\] (9)
\[
\text{var}[B(\tau_j)] = \int_0^\infty (x^2) dG^{n_j}(x^2) + [\int_0^\infty (x) dG^{n_j}(x)]^2.
\] (10)

Note that \( G^{N(\tau)}(t) \) is the \((n_1 + n_2)\)th convolution of \( G(t) \). \( G(t) \) is the cumulative density function of the boarding time for each passenger, and it follows the Normal distribution \( B_{\text{per}} \sim N(\mu, \sigma^2) \), hence the \((n_1 + n_2)\)th convolution of \( B_{\text{per}} \) still follows Normal distribution with mean and variance \((n_1 + n_2)\mu\) and \((n_1 + n_2)\sigma^2\).

Taking this property, the total passenger boarding time at stop \( s \) of line \( l \) follows Normal distribution: \( B'_l(\tau) \sim N[(n_1 + n_2)\mu, (n_1 + n_2)\sigma^2] \). If there is no vehicle bunching and delay at stops, the vehicle dwelling time at each stop is simply the passenger boarding time.

### 2.4 Modeling Passengers’ Travel Decisions:

Consider the general kind of passengers: they have a desired arrival time, but will make the travel decision with \( \alpha \) (percent) of the confidence for arrival on time. They will choose the optimal departure time and transit path as long as the \( \alpha \) (percent) confidence of on-time arrival is met. Passengers in the transit network may have different confidence levels required for their on-time arrival. Mathematically, when the stochastic user equilibrium (SUE) is reached, the optimal departure time and transit route choice problem is:
\[
\text{Min } t
\] s.t. \( P[T \leq t] \geq \alpha \) (11)

Where \( T \) is the stochastic travel time and \( \alpha \) is the stochastic confidence level that different passengers hold. Equivalently, the chance-constrained condition can be written as (the assumption of Normally distributed path travel time is hold):
\[
t \geq E[T] + \Phi^{-1}(\alpha)\text{Std}[T].
\] (12)

The last item in the above equation is also interpreted as safety margin in the literature when \( \Phi^{-1}(\alpha) > 0 \).

In the transit network, the travel cost consists of (i) passengers waiting time, (ii) in-vehicle
travel time, (iii) in-vehicle waiting time, (iii) transfer time (if transfer is needed), and (iv) the early and late arrival penalty at destination. \(ett\) is the effective travel time and defined mathematically as below. It consists of the mean of travel time and the safety margin which measures the variability of travel time.

\[
\text{ett} = E(C) + \phi^{-1}(\alpha) \cdot \text{var}(C),
\]

where \(\alpha\) is the stochastic confidence level that different passengers hold for their on-time arrival requirement. And the total generalized travel cost \(C\) is the summation of the passengers waiting time, in-vehicle travel time, the in-vehicle waiting time, transfer time:

\[
C = W(t) + \sum V_e + \sum Wv_i + [\theta + W(t')].
\]

Denote the perception error as \(\varepsilon\), which is a stochastic variable following Normal distribution. Hence the total perceived travel cost of passenger departing at time \(t\), choosing transit route \(r\) is:

\[
tc = \text{ett}_r(t) + tp(t) + \varepsilon
\]

Denote \(L\) as the vectors of passenger loads on route \(r\) between OD pair departing at time \(t\), which satisfies:

\[
\sum_{i} \sum_{r} E(L) - \sum_{i} \sum_{r} E(L') = E(Q),
\]

where \(L \in \mathbb{L}\), \(L' \in \mathbb{L}^*\), \(L^*\) is passenger alight at each transit route. Denote the maximum passenger load is \(L^a_{\max}\), the set of feasible passenger flow is \(\Omega = \{L^a \mid 0 \leq L^a \leq L^a_{\max}\}\). Then the following fixed-point problem can be derived for the proposed dynamic transit assignment model under uncertainty. In the proposed model, both the route and departure time choices are considered simultaneously:

\[
1 - q \cdot P(l) = 0
\]

where \(q\) is the vector of expected passenger O-D demand, \(l\) is the vector of mean of passenger load; and \(P\) is the vector of passenger departure time and route choice probabilities: \(l = qP(l)\).

**Theorem.** At least one solution of the fix-point problem exists.

**Proof.** The \(\Omega\) is a convex and compact set, and \(P(l)\) is continuous on \(\Omega\), then follows the Fixed Point Theorem (24), at least one solution exists for the above fixed point problem.

### 3. UPPER LEVEL MODEL: INTEGRATED OPTIMIZATION OF TRANSIT EFFICIENCY AND RELIABILITY

For the frequent transit users, assuming they know well of the transit vehicle arrival feature, they will adapt the travel decision to minimize their travel cost, especially when in a congested transit network, where the long waiting time and overload delay often affect passengers’ travel decision. The transit scheduling model considering the passenger demand variation within the investigating period. The demand and service interaction in a congested network with demand and supply uncertainties is considered by formulating a bi-level model. The transit headway, as the decision variable of the upper level model and the input data of the lower level model, is not necessarily even.

In general, network design problem are concerned with two groups: network planners and network users. On the one hand the behavior of network users follows the user-equilibrium principle of Wardrop. On the other hand network planners try to minimize the total cost of the
network. According to transit network characteristics, a bi-level programming model with a RSUE sub-model is structured as follows:

\[
(U-L) \quad \min Z = \gamma_1 \text{E}[TC \cdot P(l,h_n)] + \gamma_2 \text{var}[TC \cdot P(l,h_n)]
\]

where \( P(l,h) \) solves:

\[
(L-L) \quad 1 - q \cdot P(l,h) = 0
\]

which is given by Eq. (17). \( h \) is the vector of headways which constitute the connection of the upper and lower level models. \( h \) is assumed to be integers and the feasible lower and upper constraints for the headways of each line is held. \( \gamma_1 \) and \( \gamma_2 \) is the parameters representing the designers value on network travel time efficiency or reliability. These two parameters can be adjusted to meet the evaluation demand of different transit networks.

4. GENETIC ALGORITHM

The formulated bi-level problem is solved by a Genetic Algorithm (GA), which is altered with some constraints and intelligence. The lower level is a Probit-based dynamic assignment model, and is solved for the generation of each feasible headways by the MSA-type of algorithm. The genetic algorithm for solving the bi-level problem is outlined as follows:

**Step 0** At the initial generation of the GA, a number of populations are produced.

**Step 1** Perform the reliability-based dynamic transit assignment with given feasible transit route set between each OD pair:

*Step 1.1* Initialize the transit passenger flows on transit routes and departure time;

*Step 1.2* Simulate the PAB process and obtain the passenger flows based on the current transit travel costs and system attributes, using the Monte Carlo simulation;

*Step 1.3* Update the transit passenger flows of the two classes using the method of successive averages (MSA);

*Step 1.4* Check the convergence of the inner iteration, and get the mean and variance of the network travel time.

**Step 2** Choose the best two designs according to the fitness function from the last generation (the parents), simply the objective function, are kept in set.

**Step 3** Crossover and mutation points are randomly chosen for the evolution of the next generation.

**Step 4** Check the stopping criteria of the outer iteration.

5. NUMERICAL EXAMPLE

The small transit network as shown in Figure 2 is used for the numerical test which is similar to the test network adopted in Lam *et al.* (1999) and De Cea and Fernandez (1993). However, we slightly redefine the transit services in which the Line 4 is deleted, and the express service of Line 1 directly linking Node 1 to Node 4 is altered to link every node between them. This alternation is to better represent service lines in a CBD area for modeling morning or evening peak travel condition.

The numerical example will show the result of schedule design in the network with three lines connecting three OD pairs. The morning peak period between 8:00-9:00 is considered. The routes associated with OD pairs by the lines, segments, and transfer nodes are shown in Table 2. It is assumed that the maximal number of transfer is 1.
Some input data are: the early or late penalty $\beta^* = 0.5$, $\beta^* = 2$; the vehicle capacity for Line1, Line2 and Line3 are the same: $cap_{L1} = cap_{L2} = cap_{L3} = 120$; the original headway of each line is $oh_{L1} = 7s$, $oh_{L2} = 8s$, $oh_{L3} = 10s$; and The mean of OD demand is $od1 = 400$, $od2 = 600$, $od3 = 200$.

Table 3 shows the service attributes comparison of three lines before and after the transit scheduling design of integrated value of efficiency and reliability ($\gamma_1 = 0.5$, $\gamma_2 = 0.5$) with different OD multiplier. It can be seen, the reasonable transit scheduling can not only saving the passengers’ total travel time and enhancing service reliability, but also can reschedule vehicles to other lines to alleviate the congestion. For the normal (OD multiplier: 1.2) and congested (OD multiplier: 1.2) condition, both integrated passengers’ travel cost and cost of uncertainties have decreased after the schedule design. It is also shown that one vehicle in Line 2 is scheduled to Line 1 when demand increasing.

In order to better understand the need of rescheduling one vehicle from Line 2 to Line 1, vehicles’ occupied and available capacity before and after the schedule design are illustrated in Figure 3. When the network is not very congested, utility of vehicles has been improved, but not much. Two vehicles in Line 2 are not utilized at all before and after the schedule design. This could be explained as the number of vehicles is enough to accommodate passenger demand. However, when demand increased by 20%, a shortage of vehicle of Line 1 could be observed. As a result of the schedule design, one vehicle is scheduled from Line 2 to Line 1 to mitigate the shortage. Besides, one vehicle in Line 2 remains unutilized, but not rescheduled to any other lines. This is the consequence of the constraint of maximum headway (thus the fleet size) of each line.
Table 3 Service attributes comparison before and after the optimal transit schedule  
\(^{(\gamma_1 = 0.5, \gamma_2 = 0.5)}\)

<table>
<thead>
<tr>
<th>OD multiplier</th>
<th>1.0</th>
<th>1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispatching headways (min)</td>
<td>Even</td>
<td>Balanced</td>
</tr>
<tr>
<td>Line 1</td>
<td>8</td>
<td>11,8,9,10,6,5,9</td>
</tr>
<tr>
<td>Line 2</td>
<td>8</td>
<td>10,9,5,5,9,4,10</td>
</tr>
<tr>
<td>Line 3</td>
<td>14</td>
<td>15, 10, 11, 20</td>
</tr>
<tr>
<td>Integrated passengers travel cost</td>
<td>9614.5</td>
<td>9043</td>
</tr>
<tr>
<td>Cost of uncertainties</td>
<td>2433.5</td>
<td>2207</td>
</tr>
</tbody>
</table>

The change of vehicle schedules will impact passengers’ travel decision. It can be seen from Figure 4, The passenger demand on of OD 1 on Route 4 is more smooth after the scheduling design of transit network. The more discretized passenger distribution is intuitively benefit for saving passenger waiting time and travel time, and maintaining transit service reliability.
6. CONCLUSIONS

This paper proposed a new bi-level model for solving the transit scheduling problem in dynamic and stochastic transit network. The upper level is to change the transit scheduling route so as to optimize the transit network efficiency and reliability. The lower level is a dynamic transit assignment model which explicitly considered the demand and supply uncertainties, as well as passengers’ behavioral responses. The bi-level model was solved by GA, which could produce a stabled approximation of the optimal transit schedule by route.

The numerical example showed the result of the transit scheduling model and the performance of the solution algorithm. The data of service configuration after the scheduling design shows the saving of integrated network travel cost, while the number of vehicles between different lines is also balanced. The passenger loading profile after the alternation of vehicle schedule was smoothed, which implies better transit level of service.

ACKNOWLEDGEMENTS

The work described in this paper was jointly supported by two grants from the Research Grant Council of the Hong Kong Special Administrative Region to the Hong Kong Polytechnic University (Project Nos. PolyU 5184/05E, PolyU 5261/07E and PolyU 5202/06E). The first author would like to thank The Hong Kong Polytechnic University for support to carry out her MPhil research under the supervision of the second and third authors in Hong Kong.

REFERENCES


