J. Cent. South Univ. (2012) 19: 3603–3613
DOI: 10.1007/s11771-012-1448-9

Springer

A simulation model for estimating train and passenger delays in large-scale rail transit networks

JIANG Zhi-bin(江志彬)[1], LI Feng(李锋)[1], XU Rui-hua(徐瑞华)[1], GAO Peng(高鹏)[2]

1. School of Transportation Engineering, Tongji University, Shanghai 201804, China;
2. IBM China Research Lab, Beijing 100094, China

Abstract: A simulation model was proposed to investigate the relationship between train delays and passenger delays and to predict the dynamic passenger distribution in a large-scale rail transit network. It was assumed that the time varying original-destination demand and passenger path choice probability were given. Passengers were assumed not to change their destinations and travel paths after delay occurs. Capacity constraints of train and queue rules of alighting and boarding were taken into account. By using the time-driven simulation, the states of passengers, trains and other facilities in the network were updated every time step. The proposed methodology was also tested in a real network, for demonstration. The results reveal that short train delay does not necessarily result in passenger delays, while, on the contrary, some passengers may get benefits from the short delay. However, large initial train delay may result in not only knock-on train and passenger delays along the same line, but also the passenger delays across the entire rail transit network.

Key words: delay simulation; passenger delay; train delay; rail transit network; timetable

1 Introduction

Unexpected primary delays arise frequently in rail transit systems due to failure of equipment, signal malfunctions or varying passenger boarding or alighting times. Once the buffer time between a delayed train and the successive is smaller than the primary delay, the delayed train will hinder following trains by still occupying routing and preventing other trains from passing. It thus causes knock-on (or second) delays to other trains along the same line. Train delay is mainly a function of the primary delay time, the delay adjustment strategy and the buffer time in sections or stations.

Passenger delay is defined as the difference between actual arrival time and planned arrival time at the destination. Passenger delay is generated by the train delays, but it differs from train delays [1]. Since rail transit network has the characteristics of little headway, simply track layout and large passenger flow, a delay in one line may cause additional passenger delays on the other intersecting lines by the action of transferring, boarding and alighting, and it would take a long time to restore from disturbances.

But not all passengers may cause passenger delays after train delay, because passengers always select the first arrival train after entering in the station platform, instead of the planned train. Some passengers may arrive at destinations before the scheduled time because of less waiting time or in-vehicle running time in the process of delay propagation. Therefore, the estimation of passenger delays caused by knock-on train delays should receive high attentions.

Due to the differences in network structure, vehicle turnaround, ticket sale mode and passenger travel behavior of passengers between urban rail transit and railway, delay propagation and impact in urban rail transit are much different to delay propagation and impact in railway, which mainly reflect in the following aspects: 1) As to railway, the same track may be served by many trains running on different railway lines, and most passengers are not needed to transfer from one line to another. But in rail transit trains can only run along one line, and passengers always need transfer between transit lines to reach the final destination; 2) Due to passengers have booked the train and seat before their trips by railway, train delay inevitably leads to passenger delays. But the tickets of urban rail transit do not correspond to specific train number, and passengers can choose train service whenever necessary, thus some passengers may take a better connection caused by a train delay than the planned; 3) Because trains on

Foundation item: Project(51008229) supported by the National Natural Science Foundation of China; Project supported by Key Laboratory of Road and Traffic Engineering of Tongji University, China
Received date: 2011−11−09**; Accepted date:** 2012−02−29
Corresponding author: LI Feng, PhD Candidate; Tel: +86−21−69583771; E-mail: mapleaf2008@163.com

different railway lines always share a common track, a delay to one train can cause knock-on delays to trains running on multiple railway lines. In rail transit, train delay does not have a direct effect on trains running on other lines, but may indirectly cause knock-on delays to trains running on other lines by prolonging passenger boarding and alighting time if primary delay is long enough.

Therefore, it is unable to estimate passengers delay in rail transit using delay calculation of railway. In order to estimate train and passenger delays in rail transit network with different initial delays, it is imperative to have a delay estimation technique that is capable of accurately predicting the movement of each train and each person in rail transit networks. In the past, researchers have used either analytical methods or simulation-based methods to assess delays in railway networks or rail transit networks.

1.1 Review on analytical methods

With respect to train delays and passenger delays, the most research work focused on the train delays in railway or rail transit. CAREY and KWIECINSKI [2] focused on recovery time in their stochastic analysis. CAREY [3] used heuristic measures for timetable reliability. Furthermore, he concluded the behavioral response of drivers in some models [4]. These models have given a good insight into delay propagations on one line or a simple network. However, it becomes too complicated to handle when dealing with large scale real world networks. HALLOWELL and HARKER [5] presented an analytical line delay model that predicted the expected delay caused by a certain meet/pass plan. HIGGINS and KOZAN [6] developed an analytical model to quantify the expected delay for individual passenger trains in an urban rail network. HUISMAN and BOUCHERIE [7] provided a stochastic model to forecast secondary delays due to differences in speed of trains on railway sections. They summarized the key factors influencing running time, including number of trains, heterogeneity, primary delay, train order and buffer time. YUAN and HANSEN [8] proposed an analytical stochastic model for estimating the propagation of train delays in stations, taking the knock-on delays caused by route conflicts and late transfer connection into account. MEESTER and MUNS [9] developed a stochastic model for delay propagation and that in a word of so-called phase-type distributions. It is possible to derive secondary delay distributions from primary delay distributions. GOVERDE [10] described a delay propagation algorithm based on a timed event graph representation of a scheduled railway system. This model took into account running time supplements and dwell buffer times to recover from delays and buffer times to reduce delay propagation to connecting or conflicting trains.

1.2 Review on simulation models and tools

Simulation techniques can be used to study direct, knock-on and compound train or passenger delays and ripple effects from conflicts at complex junctions, terminals, and railroad crossings. MURALI et al [11] presented a simulation-based technique to generate delay estimates over track segments as a function of traffic conditions and the network topology. NIELSEN et al [1] presented and discussed different methods and models (“0−3 generation”) to calculate passenger delays. The evaluation of passenger delays obtained with simulation software (RailSys) and the passenger punctuality model were compared to the daily operation of the Copenhagen Suburban Network. Other delay simulation software tools, such as SIMON and Open Track, were widely used in railroad network simulations [12], which are mainly used to optimize network and timetable design. In rail transit line, JIANG et al [13] brought forward a multi-agent delay simulation model based upon the train agent-gather, adjusted strategy agent-gather, and simulation environment. This model was applied to the Line 3 and Line 4 of rail transit in Shanghai City, China.

A new simulation model is presented in this work to assess the rail transit network train and passenger delay in different conditions. Most of the existing studies emphasize on train delay evaluations in railway lines or networks. Little attention has been paid on passenger delay evaluation in large-scale rail transit networks. Train delay propagation is a complicated process, which depends on not only the relationship between trains and infrastructures, but also passengers’ reaction to delays which is a dynamic process. It is difficult to use analytical models to evaluate train and passenger delays in large-scale rail transit networks. Since computer simulation could offer rather detailed representation of a rail system, more realistic rules of trains and travel behavior, it is very beneficial to give different adjustment strategies to predict the rescheduling results. Therefore, it should be the most appropriate and reasonable way to model train and passenger delays.

Given the time varying origin-destination (OD) trip demand, scheduled timetable and realized timetable, a simulation-based approach is developed to deal with the train and passenger delay problems in this work. All transit vehicles have a limited capacity and operate precisely as specified in the scheduled timetable. Passengers queues at platforms are according to the single channel first-in-first-serve (FIFS) discipline.

2 Model formulation

All activities of trains and passengers over rail transit networks, including train running, stopping and passenger transferring, depend on the transit physical network, the scheduled timetable, queue rules and other constraints. In this section, some related models are established.

2.1 Transit physical network definition

The transit physical network is the foundation of train operation and passengers' path choices. Let $G=(S,E,L)$ be an undirected graph representing the physical network, where $S=\{s_{11}, s_{12}, \ldots, s_{ij}, \ldots, s_{mn}\}$ denotes nodes representing stations, $E=\{e_1, e_2, \ldots, e_w\}$ is the set of edges representing sections, and $L=\{l_1, l_2, \ldots, l_m\}$ is the set of transit lines in the network. Let $s^{\mathrm{T}}\{s_{i_1j_1}, s_{i_2j_2}, \mathrm{L}\}$ represents the transfer station, so the set of transfer stations can be represented as $s^{\mathrm{T}}=\{s_1^{\mathrm{T}}\{s_{i_1j_1}, s_{i_2j_2}, \mathrm{L}\}, s_2^{\mathrm{T}}\{s_{i_3j_3}, s_{i_4j_4}, \ldots\}, \ldots\}$. The presentation of a transit physical network is shown in Fig. 1 (a).

For the sake of describing the process of queuing and transferring of passengers within a station, a more detailed model of transfer stations is established, as illustrated in Fig. 1 (b). A transit transfer station is represented by a number of interconnected nodes, with some nodes representing the fare gates, some nodes representing the platforms, and other nodes representing trains. The links between fare gates and platforms represent entry and exit channels, and the links between different platforms represent the transfer channels.

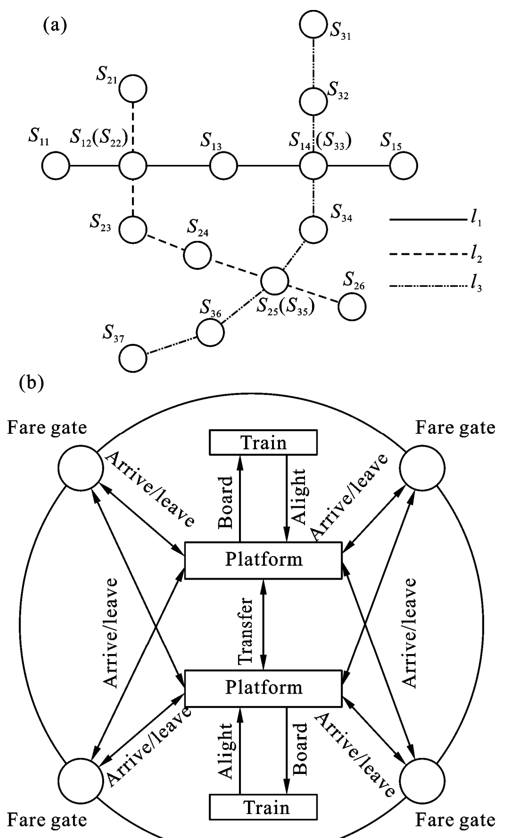

Fig. 1 Examples of rail transit physical network (a) and transfer stations (b)

2.2 Scheduled and realized timetables

Train operation network is composed by trains running following the scheduled timetable. Timetable is the basis of train operation simulation. For the purpose of estimating the passenger delays precisely, the simulation model includes two timetables: one is scheduled (planned) timetable and the other is realized timetable after an initial delay. The former is an operational schedule without any disturbance. All transit vehicles operate precisely according to scheduled timetable. The latter is simulated schedule which reflects the actual arrival and departure times of the trains affected by disruption. Figure 2 shows the scheduled timetable and the realized timetable. The location, duration and direction of the initial delay are also shown in Fig. 2. The realized timetable is produced based on the following rescheduling strategy.

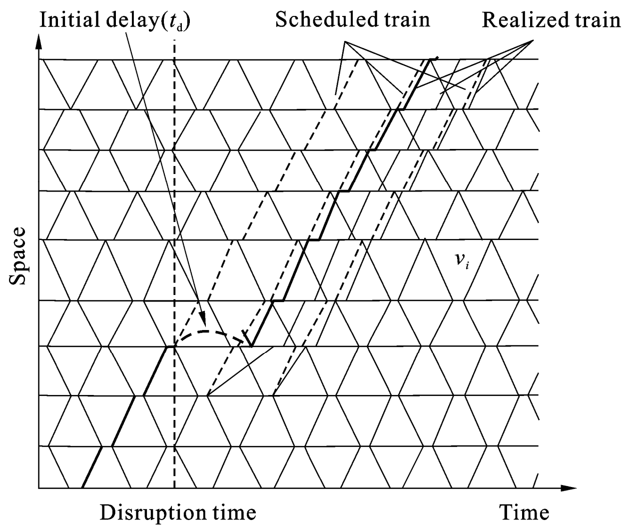

Fig. 2 Scheduled and realized time-space diagram

1) In all stations, the occupation order of trains remains unchanged;

2) The train arriving at the station ahead of schedule should also depart from the station according to the scheduled timetable;

3) The realized train timetable is according to the buffer time proportion of dwell time, inter-station running time or turnaround time. The buffer time ratio can be calculated as [14]

$$r=\frac{t_{\mathrm{b}}}{T} \tag{1}$$

where t_{b} is buffer time, and T is dwell time (or section running time, or turnaround time).

4) Delay adjustment will not stop until all trains have restored to the defined time of scheduled timetable.

The physical network has a fleet of trains V, each $v\subseteq V$ with a maximum capacity C_v. Set $a^{\mathrm{S}}_{v,s}$ as scheduled arrival time of train v at station s, $a^{\mathrm{R}}_{v,s}$ as realized arrival time of train v at station s, and $d^{\mathrm{S}}_{v,s}$ as scheduled departure time of train v at station s and $d^{\mathrm{R}}_{v,s}$ as realized departure time of train v at station s. The scheduled timetable and realized timetable can be respectively expressed as

$$\begin{cases} T^{\mathrm{S}}=\{s_{i,j}(a^{\mathrm{S}}_{v_{i,k},s_{i,j}},d^{\mathrm{S}}_{v_{i,k},s_{i,j}})\} \\ T^{\mathrm{R}}=\{s_{i,j}(a^{\mathrm{R}}_{v_{i,k},s_{i,j}},d^{\mathrm{R}}_{v_{i,k},s_{i,j}})\} \end{cases} \tag{2}$$

where T^{S} is scheduled timetable, $T^{\mathrm{R}}_{s,t_{\mathrm{d}}}$ is realized timetable aiming to delay occurring at station s, and initial delay time is t_{d}.

2.3 OD demand and passenger path choice

The size and distribution of network passenger flow are the basis of transit scheduling and passenger flow simulation. Since the AFC system could accurately record historical data of passengers' origins, destinations, and entry and exit times, the time-varying demand and path choices can be derived from it.

Let O and D represent the sets of the origins and destinations, respectively; R^{OD} represents the set of valid paths from O to D, and $R^{OD}=\{r^{OD}_1, r^{OD}_2, \cdots r^{OD}_i\}$. Set P as the total flow entering the rail transit. Then $P(O, D, t^P_O, r^{OD})$ means the passengers who enter station O at time t^P_O, select the path r^{OD}, and leave the rail transit system from station D finally.

In a large-scale urban rail transit network, several feasible paths always exist between a pair of origin-destination, but not all are valid. Valid paths are part of the feasible paths which get rid of some unreasonable paths, such as outflanking paths, and long travel time paths. Factors, such as the distance, in-vehicle time, transfer time and number of transfers, are different, thus the costs of different paths are different. Define the $\omega_{r^{OD}}$ as the probability of passengers choosing path r^{OD}. It holds that

$$\sum_{r^{OD}\in R^{OD}}\omega_{r^{OD}}=1 \tag{3}$$

The table of $\omega_{r^{OD}}$ can be obtained from the AFC system in Shanghai City. It must be noted that our model assumes that the destination and path choice of each passenger are not affected by failures, that is, all passengers will keep travelling to their planned destinations regardless of disruptions in the normal functioning of the network.

2.4 Passenger travel time calculation with capacity constraints

Travel time of passengers in the rail transit includes four components:

1) t^{E}_O : Entry walking time from the fare gate to the platform at station O;

2) $t^{\mathrm{W}}_{p,s}$: Waiting time before boarding on a train for passenger at station O or a transfer station;

3) $t^{\mathrm{I}}_{p,v}$: In-vehicle time from boarding to alighting on train v;

4) $t^{\mathrm{T}}_{s^{\mathrm{T}}}$: Transfer time at station S^{T}.

So, the time when passenger p is alighting from the train v at station D can be expressed as

$$T^p_D=T^p_O+t^{\mathrm{E}}_O+\sum t^{\mathrm{W}}_{p,s}+\sum t^{\mathrm{I}}_{p,v}+\sum t^{\mathrm{T}}_{s^{\mathrm{T}}} \tag{4}$$

where T^P_O is the time when passenger p enters station O.

The whole travel time from entering at station O to alighting from station D for passenger p is

$$t_{P(O,D,t^P_O,r^{OD})}=t^{\mathrm{E}}_O+\sum t^{\mathrm{W}}_{p,s}+\sum t^{\mathrm{I}}_{p,v}+\sum t^{\mathrm{T}}_{s^{\mathrm{T}}} \tag{5}$$

Train delays do not necessarily cause passenger delays. Some passengers may even benefit from train delays. If a passenger arrives late to the station, a train delay may allow the passenger to catch an earlier train than expected (see p_2 in Fig. 3).

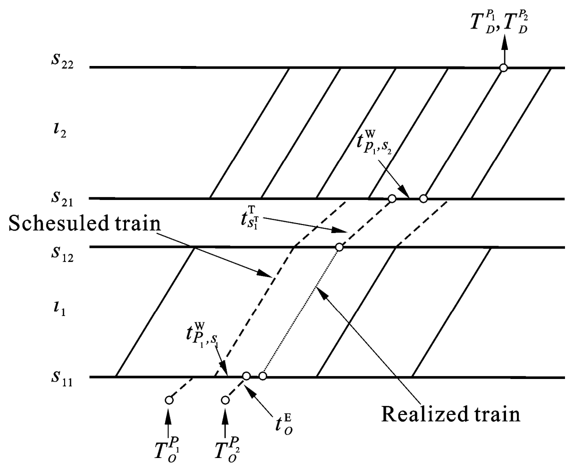

Fig. 3 Passenger travel action analysis based on scheduled and realized timetables

During the peak hours, passengers at crowded stations may need to wait for a long time because they have failed to board the first train they wait for. It is assumed that all transit vehicles have limited capacities. Two policies (i.e. unlimited serviceable capacity (UCS), limited serviceable capacity & first-in-first-server (LSC & FIFS)) are put forward to describe alighting and boarding of passengers. When a train arrives at a station, alighting procedure precedes boarding procedure. Alighting is not constrained by platform capacity, that is, passengers are able to alight regardless of the number of waiting passengers at the platform. There is no priority among alighting passengers. All of them follow the principle of equality and alight from the train simultaneously. In other words, the policy could be formulated as UCS: any passenger alighting could be served instantly. As to the process of boarding, a passenger has priority over the rear of the queue to board. The waiting time of a boarding passenger is determined by the queue length in front of him/her and the residual capacity of the train being boarded. The policy for the boarding process is FIFS with limited capacity.

Assume that L_s^v is the number of on-board passengers of arrival vehicle v at station s, W_s^v is the number of passengers waiting for train v at station s, and A_s^v is the number of passengers which need to alight from train v at station s. So, the realized number of passengers boarding in train v at station s is

$$B_s^v=\begin{cases}W_s^v, \text{ if } W_s^v-(C_v-L_e^v+A_s^v)\le 0\\ C_v-L_s^v+A_s^v, \text{ otherwise}\end{cases} \tag{6}$$

where $W_s^v-(C_v-L_s^v+A_s^v)>0$, and it means that there are $[W_s^v-(C_v-L_s^v+A_s^v)]$ passengers unable to board and need to wait for the next train. So, the waiting time of the passengers who board train at his/her original station and transfer station could be respectively calculated as

$$t_{p,s}^{\mathrm{W}}=d_{v,s}^{\mathrm{R}}-(T_O^p+t_O^{\mathrm{E}}) \tag{7}$$

$$t_{p,s}^{\mathrm{W}}=d_{v,s}^{\mathrm{R}}-(a_{v',s}^{\mathrm{R}}+t_{s^T}^{\mathrm{T}}) \tag{8}$$

where $a_{v',s}^R$ represents the realized arrival time of the train which passenger p boards before his/her transfer at station s.

2.5 Train delay and passenger delay

2.5.1 Train delay

If the actual arrival time or departure time of train v is behind schedule, which means that the train has delayed, denoted as notes for $D_j^{v_k}=1$, otherwise $D_j^{v_k}=0$. $D_j^{v_k}$ could be calculated as

$$D_{l_i}^{v_k}=\begin{cases}1 & \text{if } d_{v_k,s_{i,j}}^{\mathrm{R}}>d_{v_k,s_{i,j}}^{S} \text{ or } a_{v_k,s_{i,j}}^{\mathrm{R}}>a_{v_k,s_{i,j}}^{S},\\ & \forall v_k\in V, l_i\in L, (i,j,k)=1,2,\mathrm{L}\;, n\\ 0 & \text{else}\end{cases} \tag{9}$$

Define $n_{N,d}^V$ as the total number of delayed trains in network, and $n_{s,d}^V$ as the total number of delayed trains at station s, then

$$\begin{cases}n_{N,d}^V=\sum D_{l_i}^{v_k}, \quad \forall v_k\in V, l_i\in L, (i,k)=1,2,\cdots,n\\ n_{s,d}^V=\sum D_s^{v_k}, \quad \forall v_k\in V, s\in S, k=1,2,\cdots,n\end{cases} \tag{10}$$

$$D_s^{v_k}=\begin{cases}1 \text{ if } d_{v_k,s}^R>d_{v_k,s}^S \text{ or } a_{v_k,s}^R>a_{v_k,s}^R, \forall v_k\in V, s\in S\\ 0 \text{ else}\end{cases}$$

Define $t_{s,d}^V$ as the total train delays at station s, then

$$t_{s,d}^V=\sum[(d_{v_k,s}^{\mathrm{R}}-d_{v_k,s}^{\mathrm{S}})+(a_{v_k,s}^{\mathrm{R}}-a_{v_k,s}^{\mathrm{S}})], \ \forall v_k\in V_K \tag{11}$$

where V_K stands for the collection of trains which go through station s.

Likewise, the total train delays in the network can be calculated as

$$t_{N,d}^V=\sum t_{s,d}^V, \ \forall s\in S \tag{12}$$

2.5.2 Passenger delay

1) Passenger delay

If the passenger's travel time in the case of the scheduled timetable is regarded as a standard, the passenger's delay in the realized timetable can be expressed as

$$t_{D,d}^p=T_{p,D}^{\mathrm{R}}-T_{p,D}^{\mathrm{S}} \tag{13}$$

Define $t_{D,d}^p$ as the delay time at station D for passenger p, $T_{p,D}^{\mathrm{R}}$ as the time when alighting from the train at station D for passenger p in the case of realized timetable, $T_{p,D}^{\mathrm{S}}$ as the time when alighting from the train at station D for passenger p in the case of scheduled timetable.

According to different $t_{D,d}^p$ values, positive delays, no delay and negative delays can be defined as $t_{D,d}^p>0, t_{D,d}^p=0,$ and $t_{D,d}^p<0,$ respectively.

2) Total passenger delays of stations

The total positive delays ($t_{D,\mathrm{d}}^{\mathrm{p},p}$), negative delays ($t_{s,\mathrm{d}}^{\mathrm{n},p}$), and total passenger delays ($t_{s,d}^p$) of station s can be expressed as

$$t_{s,\mathrm{d}}^{\mathrm{p},p}=\sum t_{D,\mathrm{d}}^p \quad \forall t_{D,\mathrm{d}}^p>0, D{=}s, p\in P \tag{14}$$

$$t_{s,\mathrm{d}}^{\mathrm{n},p}=\sum t_{D,\mathrm{d}}^p \quad \forall t_{D,\mathrm{d}}^p<0, D{=}s, p\in P \tag{15}$$

$$t_{s,\mathrm{d}}^p=t_{s,\mathrm{d}}^{\mathrm{p},p}+t_{s,\mathrm{d}}^{\mathrm{n},p} \tag{16}$$

3) Total passenger delays in network

The total positive delays ($t_{N,\mathrm{d}}^{\mathrm{p},p}$), negative delays ($t_{N,\mathrm{d}}^{\mathrm{n},p}$), and total passenger delays ($t_{N,\mathrm{d}}^p$) in network can be expressed as

$$t_{N,\mathrm{d}}^{\mathrm{p},p}=\sum t_{s,\mathrm{d}}^{\mathrm{p},p} \quad \forall s\in S \tag{17}$$

$$t_{N,\mathrm{d}}^{\mathrm{n},p} = \sum t_{s,\mathrm{d}}^{\mathrm{n},p} \quad \forall s \in S \tag{18}$$

$$t_{N,\mathrm{d}}^{p} = \sum (t_{s,\mathrm{d}}^{\mathrm{p},p} + t_{s,\mathrm{d}}^{\mathrm{n},p}) \quad \forall s \in S \tag{19}$$

4) Total number of delayed passengers

The total number of delayed passengers ($n_{s,\mathrm{d}}^{p}$) and passengers ($n_{s,\mathrm{b}}^{p}$) who leave transit station s ahead of the scheduled can be expressed as

$$n_{s,\mathrm{d}}^{p} = \sum D_D^p, \forall p \in P, D = s \tag{20}$$

$$n_{s,\mathrm{b}}^{p} = \sum B_D^p, \forall p \in P, D = s \tag{21}$$

$$D_D^p = \begin{cases} 1 & \text{if } t_{D,d}^p > 0 \\ 0 & \text{else} \end{cases}$$

$$B_D^p = \begin{cases} 1 & \text{if } t_{D,d}^p < 0 \\ 0 & \text{else} \end{cases}$$

Likewise, for all passengers in the network, the total number of delayed passengers ($n_{N,\mathrm{d}}^{p}$) and the total number of passengers ($n_{N,\mathrm{b}}^{p}$) who reach their destinations ahead of the scheduled in the case of the realized timetable can be expressed as

$$n_{N,\mathrm{d}}^{p} = \sum n_{s,d}^{p}, \forall s \in S \tag{22}$$

$$n_{N,\mathrm{b}}^{p} = \sum n_{s,b}^{p}, \forall s \in S \tag{23}$$

3 Assumptions and solution procedure

3.1 Assumptions and limitation

The following assumptions and limitations are made for the problem:

1) Origin–destination demand matrix and entry time of each passenger are known;

2) Passenger boarding is constrained by a fixed capacity of the vehicle being boarded;

3) Passenger alighting is not constrained by the platform capacity;

4) The safe separation between trains is ensured by imposing minimum intervals between two successive train arrivals or departures at each station. This interval is determined by the signal system;

5) The initial delay duration is limited and can be determined;

6) The OD demand will not be affected by the disruptions and nobody will change their path choice;

7) The order and route of transit vehicles are not changed in the process of rescheduling timetable;

8) The realized timetable is given, and the adjustment strategy is to shorten the dwell time and inter-station running time (the proportion of buffer time is given).

3.2 Data requirements

The following set of data is required as input to the model.

1) The physical transit network;

2) The scheduled timetable;

3) Origin–destination and entry time of each passenger;

4) The travel times of entry channels, exit channels and transfer channels at each station;

5) The capacities of vehicles and platforms;

6) Minimum intervals between two successive train arrivals or departures;

7) Disruption location and estimated duration;

8) Realized timetable after different initial delay disruptions.

3.3 Solution procedure

The main steps of the solution procedure are described as follows:

Step 1: Initialize network. Establish models of the network structure, lines, station, platform and trains.

Step 2: Load scheduled timetable and realized timetable. Input the scheduled timetable and the realized timetable in different initial delays (generated by delay recovery strategy).

Step 3: Load network passenger flow. Load a time-space trip matrix of rail transit passenger flow OD distribution (the time interval is 5 min), and then load the table of the probabilities of passengers' choices on various paths ($\omega_{r,OD}$).

Step 4: Simulation of the operation. Set running time, and use the planned timetable and the delayed timetable to simulate respectively. Dynamically display and record the information of train operation and the dynamic distribution of passenger flows. Record the entry time, transfer time and exit time of each passenger in the whole travel process.

Step 5: Calculation of train and passenger delays. Calculate the passenger travel time in two situations according to the two simulation records. By comparing the same passenger's $p \in P$ travel time in two situations, the train delays or passenger delays are calculated, return to **Step 2.**

Step 6: Evaluation. According to the train and passenger delay calculations, the range and degree of influence that the disruption exerts on rail transit network and stations can be evaluated.

4 Computational experiments

According to the proposed models, the assumption and solution procedure, a simulation software named URT_PDSS is developed (urban rail transit passenger delay simulation system). On the basis of constructing the physical rail transit network and operation network, the dynamic passenger flow distribution, calculation, simulation and statistical analysis in the normal and delayed conditions can be realized. Figure 4 shows the structure of the URT_PDSS.

Figure 5 shows the dynamic rail traffic network passenger flow distribution in the simulation process. The passenger number at each station and each operation train, inner station hall and platforms can all be checked in detail. Different color density indicate different passenger flow densities, which intuitively show the passenger service level of different areas and nodes.

4.1 Case

The framework and models are demonstrated for the case of the Shanghai Rail Transit Network of China in 2009, which has 8 rail transit lines, with 170 stations and a total length is 250.2 km (time to June 30, 2009), accommodating almost 4×10^6 passengers on a daily operation.

URT_PDSS is operated on an Intel Xeon Quad PC (3.0 GHz) with 3 GB of memory to calculate the train delays and passenger delays as formulated in Section 4. Several scenarios are simulated by URT_PDSS to estimate passenger delays caused by train delays. In this case study, it is assumed that the primary delay occurs at Shanghai Science & Technology Museum at 8:00 a. m. in downward direction, and the primary delay duration is set from 1 to 10 min. The stations of initial delay occurring and linking are shown in Fig. 6. The input OD data is actual passenger flow data of Shanghai Rail Transit Network (March 17, 2009, Tuesday). Scheduled timetable are actual working day operation timetable in March, 2009, and realized timetable can be generated by TPM [15]. The simulation horizon is selected to cover the morning peak period (7:00−11:30 a.m.) with the first half an hour as warm up period and the last hour as the cleanup period. The total passenger number is 421 763. The simulation process lasts about 1 h.

4.2 Computational results

4.2.1 Distribution of station passenger delays in rail transit network

Because the headway of trains in urban rail transit is very small, in order to maintain the stability of timetables in the case of disruption, it is not necessary to adjust timetables of other transit lines for passengers connection at transfer station, thus the influence of delays on the trains operated on other lines is relatively limited. However,

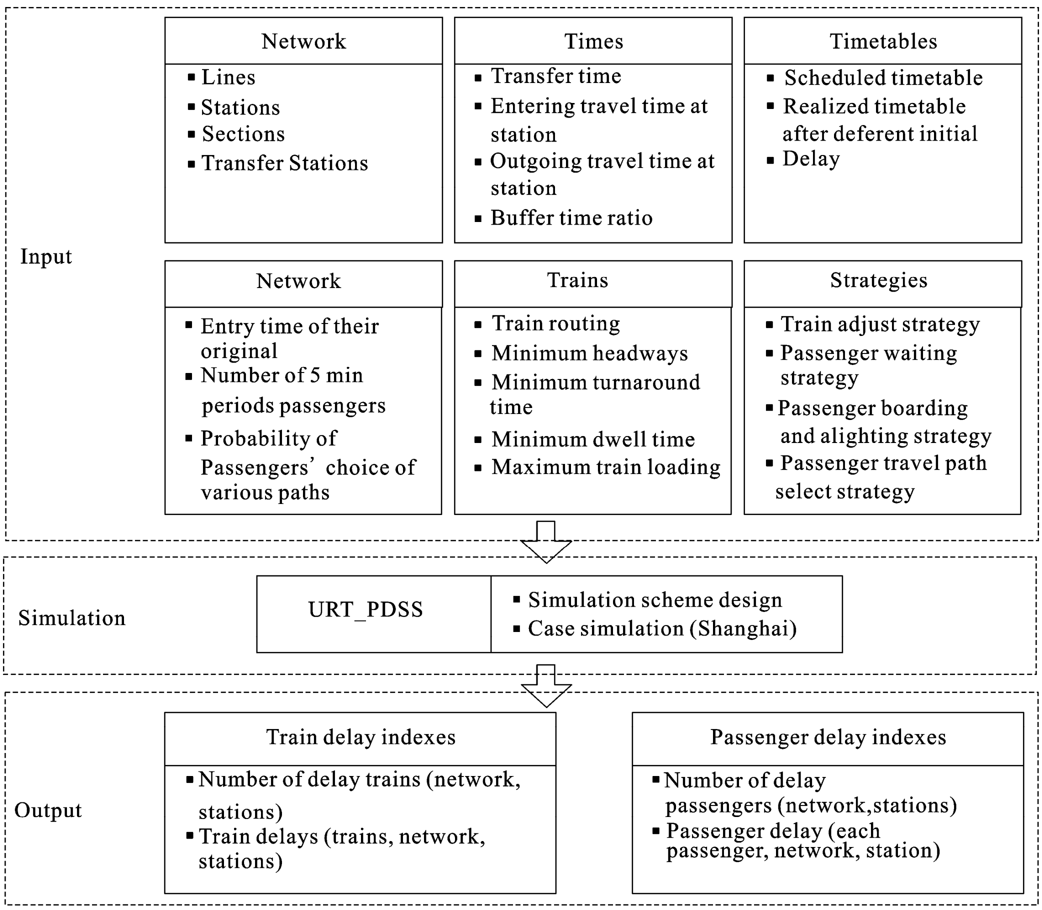

Fig. 4 Structure of URT_PDSS

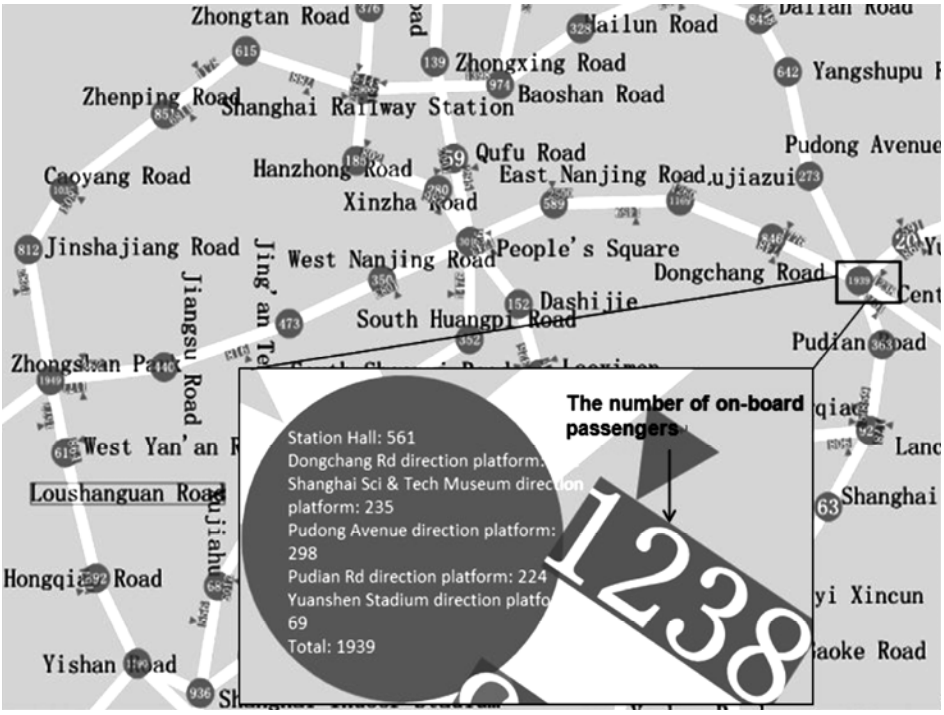

Fig. 5 Simulation interface of URT_PDSS system (dynamic distribution of passenger flow)

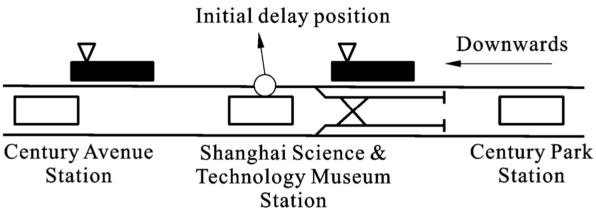

Fig. 6 Stations of initial delay occurring and linking

the effect of the disruption on passengers may cover the whole network. If only passenger passes through the delayed line (including arrival and departure), he/she is likely to be affected. Therefore, passenger delays will be distributed to the whole network. Figure 7 shows the network passenger delay's scatter plot distribution (including positive delays and negative delays) when initial delay time is 3 min and 5 min, respectively. It can be seen in Fig. 7 that, when the initial delay is small, the affected stations are less and the total passenger delay of the stations (including the positive and negative) is short, conversely, more stations are affected stations and the total passenger delays is longer.

4.2.2 Network train and passenger delay characteristics with different initial delays

Table 2, Figs. 8 and 9 represent the delay indexes of network trains and passengers with different initial delay conditions (1, 2, 3, ... , 10 min). According to Fig. 8, both the number and delays of trains present a direct proportion growth relationship to the initial delays, and the train delays tend to change exponentially, while the number of delayed trains appears linear change trend. When the initial delay time is more than 5 min, the total

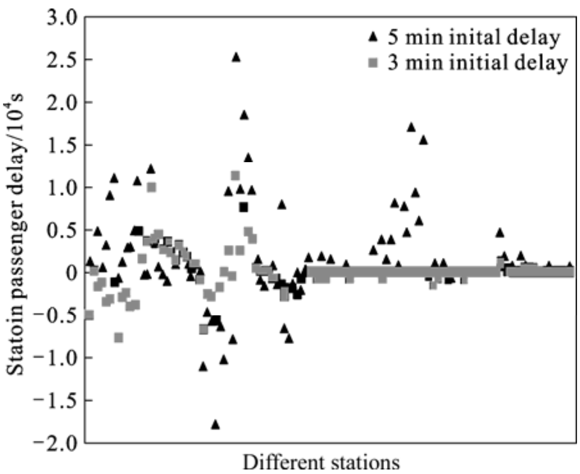

Fig. 7 Passenger delay time at each station with deferent initial delays

delay variation begins to increase.

In Fig. 9, it can be seen that train delays will not necessarily cause passenger delays. Some passengers may benefit from delays (exiting destination in advance). The total number and delays of the delayed passengers, the number of passengers and the total delays exiting at destination in advance all increase consistently with the increasing initial delays. But the number of passengers who exit in advance is always smaller than the number of delayed passengers. When the train initial delay is within 3 min, most of the passengers board ahead of scheduled time, so the time saved in the waiting process is nearly equal to the delays, the variation trend of the total positive

Table 2 Train and passenger delay index with different initial delays

Index	Initial delay time/min									
	1	2	3	4	5	6	7	8	9	10
Persons arrive in destination after scheduled	309	1455	2330	3948	4630	7646	10244	14018	17649	21213
Positive passenger delay/s	5 705	99 892	205 333	602 945	761 143	1 312 855	1 818 626	2 812 181	3 669 688	4 884 183
Persons arrive at destination ahead of scheduled	235	782	1446	2601	4193	6526	9454	13127	14556	17033
Negative passenger delay/s	42 966	128 470	203 129	384 626	569 845	1 029 748	1 352 838	1 796 844	1 945 272	2 128 259
Total passenger delay/s	−37 261	−28 578	2 204	218 319	191 298	283 107	465 788	1 015 337	1 724 416	2 755 924
Average positive passenger delay/s	18	69	88	153	164	172	178	201	208	230
Average negative passenger delay/s	183	164	140	148	136	158	143	137	134	125
Total number of delayed trains	2	4	5	6	8	11	13	15	18	20
Number of delay trains in terminal stations	0	0	1	2	3	5	6	8	10	13
Total train delay/s	350	1 824	5 167	10 592	18 412	29 307	44 033	63 158	86 722	114 672
Total arrival train delay/s	140	818	2 411	5 045	8 854	14 169	21 363	30 744	42 341	56 138
Total departure train delay/s	210	1 006	2 756	5 547	9 558	15 138	22 670	32 414	44 381	58 534

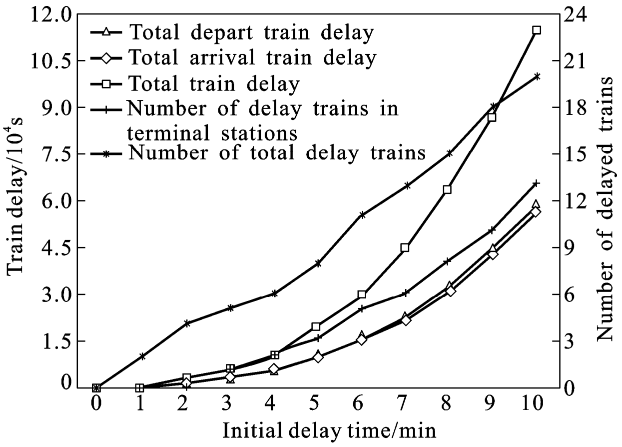

Fig. 8 Network train delay index with different initial delays

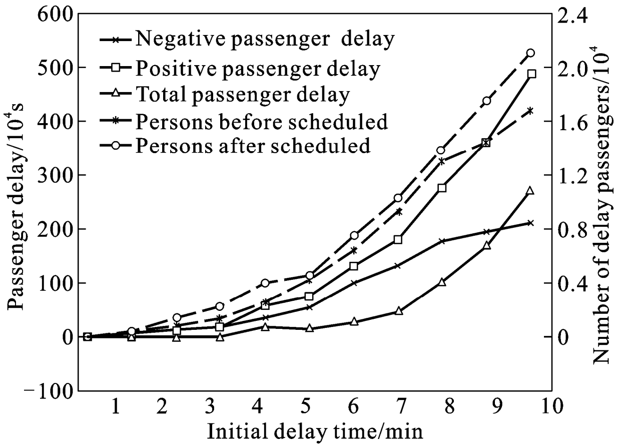

Fig. 9 Network passenger delay index with different initial delays

delays and of the total negative delays are gentle. If the initial delay is 3−6 min, the extension of travel time caused by train delays will affect most of the passengers. The total positive delay and the total negative delay grow rapidly. However, the total negative delay grows more rapidly, which offsets the total positive delay in a certain degree, thus leading to the fact that the total delay grows at a slow speed. When the initial delay is more than 7 min, positive delays grow more rapidly than the negative delays. The total delay also shows the exponential growth. Therefore, the variation trends of total delay originally appear to decrease (initial delay less than 3 min), then to be gentle (initial delay between 3 and 6 min) and then to increase (initial delay greater than 6 min) at last.

Figure 10 indicates the character of the average network passenger delays. The average positive delay and the train initial delay reflect a linear growth relationship. When the initial delay is less than 5 min, the average positive delay growth amplitude is larger; when it's more than 5 min, the growth amplitude slows down. While the average negative delay changes smoothly, and decreases gradually. This shows that as the initial delay

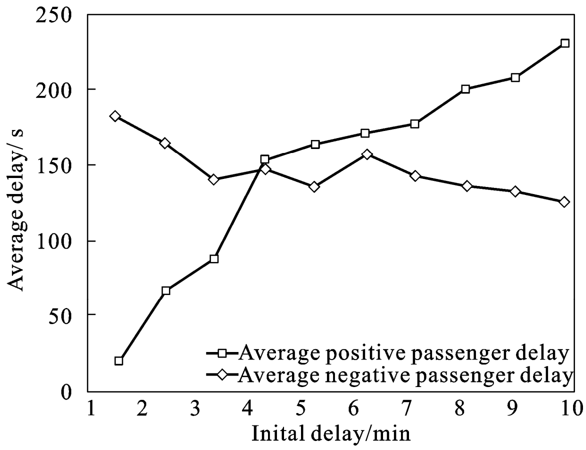

Fig. 10 Network average passenger delays with different initial delays

grows, passengers' waiting time on the platform also grows, although some passengers may catch earlier train than expected. But when the number of waiting passenger is large, which will lead to higher platform waiting time (passengers need to board on the next train), while the average negative delay appears a decreasing trend. What's more, when the initial delay is less than or equal to 4 min, the average positive delay is less than the average negative delay, which indicates that the benefit time of the passengers who exit ahead is larger.

4.2.3 Station train and passenger delay characteristics with different initial delays

Figure 11 shows the passenger and train delays at the station where initial delay occurs. According to the Fig.11, the total delay's change regularities at all the three stations are similar, which all increase with small amplitudes first (when the initial delay is less than 7 min) and then increase progressively with larger amplitudes. The change regularity is consistent with that of the network total passenger delays. While the time saved by passengers who exit in advance increases slowly at first, then increases dramatically, and drops a bit in the end. Especially at the initial delay happened station (Shanghai Science and Technology Museum Station), when the initial delay time is more than 7 min, the negative passenger delays of Shanghai Science and Technology Museum Station starts to decrease. The total passenger delay time has always been the largest of these three stations. The total passenger delay time of Century Park Station is the minimal about 1/3 of Shanghai Science and Technology Museum Station. The total train delay time of these three stations presents exponential growth. The total train delay time of Century Avenue Station is the maximum all the time, and the total passenger delay

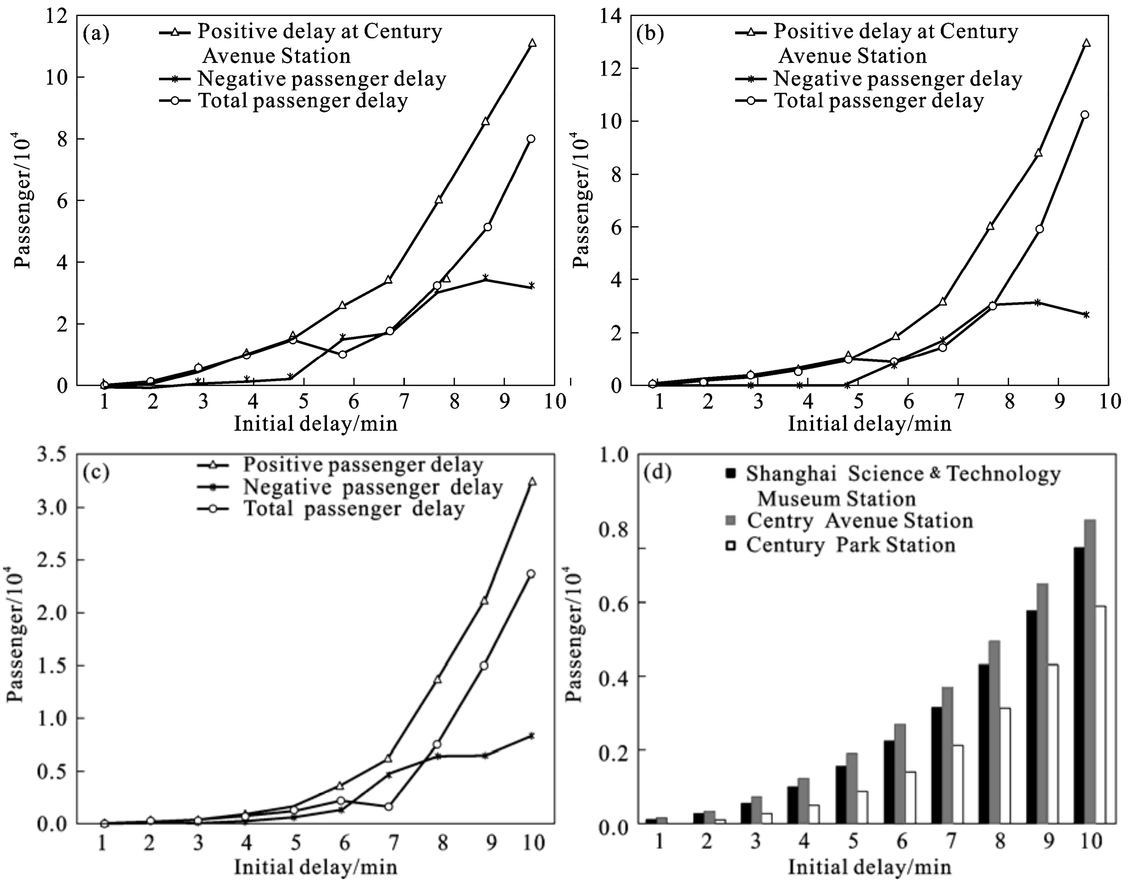

Fig. 11 Train and passenger delay at Century Avenue, Shanghai Science & Technology Museum and Century Park stations: (a) Passenger delay at Century Avenue Station; (b) Passenger delay at Century Avenue Science & technology museum Station; (c) Passenger delay at Century Park Station; (d) Train delay at different stations

time of Century Park Station is the minimal. The above analysis indicates that train delay has greatest influence on the passenger delay of the initial delay happened station, and has more influence on the train delay of the next link station. Meanwhile the influence on the latter station's passenger and train delay is smaller.

5 Conclusions and further research

1) The large initial train delays cause not only knock-on train delays in the same line, but also passenger delays in whole rail transit network. The delay number and delay time of trains both present a direct proportion growth relationship to the initial delay time, the train delay time tends to change exponentially, while the number of delayed trains appears linear change trend.

2) The small train delays will not necessarily cause passenger delays. Some passengers may benefit from delays. The total number and delay time of delayed passengers, the number of passengers exiting ahead of schedule and the total delay time are all increasingly consistent with the initial delay time. But the number of passengers who exit in advance is always smaller than the number of delay passengers. The variation trends of total passenger delay time originally appear to reduce, then to be gentle and to increase at last. The average positive delay and the train initial delay reflect a linear growth relationship, while the average negative delay changes smoothly, and gradually decreases. In addition, the influence of different stations on passenger and train delay is different.

3) Train delay has the greatest influence on the passenger delay of the station where the initial delay happened, and has the greatest influence on the train delay of the next station, while the influence on the upstream station's passenger and train delay is smaller. So, in daily operation, the smaller initial delay may not make serious effect on passengers, but more attention must be paid on passenger organization (such as, make more passengers get on the near next train), especially on the next of initial delay stations.

4) Passenger delay simulation in rail transit network is very complicated in nature, and a number of critical challenges need to be further addressed in future research. A core assumption in this work is that all passengers don't change their destinations and work choice when delay occurs. But this assumption does not reflect the fact. when delay happens, accessibility and impedance of rail transit path are changed, but passengers will not change their paths only in small delays. Train dwell time may be affected by passenger boarding or alighting. Especially in the case of over loading passengers in operation train, delays could lead to more waiting passengers on some platforms, so the delay adjustment process should consider the effect of passenger flow changes on stations.

References

[1] NIELSEN O A, LANDEX O, FREDERIKSEN R D. Passenger delay models for rail networks [M] Schedule-Based Modeling of Transportation Networks. Springer US, 2009: 27−49.

[2] CAREY M, KWIECINSKI A. Properties of expected costs and performance measures in stochastic models of scheduled transport [J]. European Journal of Operational Research, 1995, 83(1): 182−199.

[3] CAREY M. Exante heuristic measures of schedule reliability [J]. Transportation Research Part B: Methodological, 1999, 33(7): 473−494.

[4] CAREY M. Optimizing scheduled times, allowing for behavioral response [J]. Transportation Research Part B: Methodological, 1998, 32(5): 329−342.

[5] HALLOWELL S F, HARKER P T. Predicting on-time performance in scheduled railroad operations: methodology and application to train scheduling [J]. Transportation Research Part A: Policy and Practice, 1998, 32(4): 279−295.

[6] HIGGINS A, KOZAN E. Modeling train delays in urban networks [J]. Transportation Science, 1998, 32(4): 46−357.

[7] HUISMAN T, BOUCHERIE R J. Running times on railway sections with heterogeneous train traffic [J]. Transportation Research Part B: Methodological, 2001, 35(3): 271−292.

[8] YUAN J, HANSEN I A. Optimizing capacity utilization of stations by estimating knock-on train delays [J]. Transportation Research Part B: Methodological, 2007, 41(2): 202−217.

[9] MEESTER L E, MUNS S. Stochastic delay propagation in railway networks and phase-type distributions [J]. Transportation Research Part B: Methodological, 2007, 41(2): 218−230.

[10] GOVERDE R M P. A delay propagation algorithm for large-scale railway traffic networks [J]. Transportation Research Part C: Emerging Technologies, 2010, 18(3): 269−287.

[11] MURALI P. DESSOUKY M. PALMER K. MURALI P DESSOUKY M, ORDONEZ F, PALMER K. A delay estimation technique for single and double-track railroads [J]. Transportation Research Part E: Logistics and Transportation Review, 2010, 46(4): 483−495.

[12] JIANG Zhi-bin, XU Rui-hua, XIE Chao. Train delay propagation simulation in rail transit system [C]// The 1st International Conference on Transportation Engineering. Reston: American Society of Civil Engineers, 2007: 789−794.

[13] JIANG Zhi-bin, XIE Chao. Multi-agent delay simulation model in mass rail transit system [C]// 2009 International Conference on Measuring Technology and Mechatronics Automation. Los Alamitos: IEEE Computer Society, 2009: 717−720.

[14] JIANG Zhi-bin, Jin Yi, Chen Jing-jing. Running buffer time distribution calculation based on delay analysis in mass transit system [C]// The 2nd International Conference on Transportation Engineering. Reston: American Society of Civil Engineers, 2009: 4068−4073.

[15] JIANG Zhi-bin, GAO Jia, Xu Rui-hua. Circle rail transit line timetable scheduling using Rail TPM [C]// Computers in Railways XII: Computer System Design and Operation in Railways and Other Transit Systems. Southampton: WIT Press, 2010: 945−952.

(Edited by DENG Lü-xiang)