Dynamics of congested urban rail transit: Equilibrium and policy analysis

Jiahua ZHANG[1], Kentaro WADA[2] and Takashi OGUCHI[3]

[1]Non-member of JSCE, Project Researcher, Inst., Industrial Science, the University of Tokyo
(4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan)
E-mail:zhangjh@iis.u-tokyo.ac.jp
[2]Member of JSCE, Associate Professor, Faculty of Engineering, Information and Systems, University of Tsukuba
(1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan)
E-mail: wadaken@sk.tsukuba.ac.jp
[3]Fellow of JSCE, Professor, Inst., Industrial Science, the University of Tokyo
(4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan)
E-mail:takog@iis.u-tokyo.ac.jp

This study macroscopically develops a dynamic model capable of describing equilibrium distribution of passenger arrivals in a congested rail transit system based on information of timetable and passenger preferences. By applying the proposed model, insights into the design of management strategies are derived from a macroscopic view. More specifically, a conventional peak/off-peak timetable are numerically optimized under user equilibrium. The optimal peak/off-peak timetable is found to be able to significantly reduce the sum of travel time cost and schedule delay cost. Besides, a one-step coarse pricing scheme is specified and its design is numerically evaluated. It is found that temporal settings (i.e., start and end time, duration) play a crucial role in the effectiveness of the pricing scheme.

Key Words : *Equilibrium model, rail transit, timetable optimization, pricing scheme design*

1. Introduction

Urban rail transit, with its high capacity and punctuality, serves as a typical solution to commuters' travel demand during rush hours in most metropolises[1]. However, travel experience of commuting by rail transit is frequently deteriorated by severe congestion and unexpected delays. In many metropolises, congestion and delay of rail transit have brought about tremendous psychological stress to commuters and considerable economic loss to the whole society. For example, Xu et al.[2] reported that around one third of subway stations in Beijing adopted ordinary passenger flow control. Kariyazaki[3] estimated that the social cost due to the delay of trains in Japan exceeded 1.8 billion dollars per year. When a train delays due to an unexpected event (e.g., person on the tracks, signal failure or mechanical trouble), the delay propagates along the railway line. As a result, both in-vehicle crowding and on-platform congestion become severer on the whole railway line. Even no accident occurs, in a high-frequency operated urban rail transit line, chronic delays widely exist due to significantly extended dwelling time at stations or “knock-on delay” on the rail track[3),4]. Especially during the rush hours, passenger congestion on-platform and this kind of chronic delays can easily develop into a vicious circle[5]. In order to relieve congestion and prevent the occurrence of delay, management strategies have long been investigated by researchers with different backgrounds.

As an important management strategy, optimization of train timetable received considerable attention in the past decades[6) - 14]. Although most of these studies consider the time-dependent passenger demand in their optimization, the demand itself is usually treated as given information. When the timetable is modified or new pricing schemes are implemented, many passengers would probably adjust their departure time from home. Thus, the time-dependent demand itself would change, which makes the system not optimized again. Therefore, it is important to understand the interaction between passengers' decision making on their departure time and operational environment of rail transit system (e.g., timetable and congestion information, pricing scheme). As a transportation demand management (TDM) strategy, motivating commuters to change their departure time choice is considered to be effective to temporally flatten the surging demand.

Economists like Vickrey[15] and Henderson[16] might be the first to consider individuals' departure time decisions for commuting problem by private cars. They considered the existence of a bottleneck on the road and assumed that passengers minimized their total travel costs to determine their departure time. When the total travel cost is minimized for any individual, no one has the incentive to change his or her departure time, which is generally referred to as an equilibrium state. So far, many studies had thoroughly discussed the equilibrium distribution of arrivals at the bottleneck, either in a deterministic[17)–20)] or stochastic setting[21)–23)]. However, these bottleneck models for road traffic may not be readily applicable to rail transit since the mechanism of congestion and delay are quite different between these two systems. As mentioned above, the delay of trains in many densely populated Asian cities are mainly caused by the vicious circle of passenger congestion and bunching phenomenon of high-frequency operated trains. Therefore, equilibrium model targeted at the commuting problem for these cities should take into consideration of this mechanism. While in fact, most studies tackling on rail transit assumed a constant travel time[24)–29)], which implied that delay of trains due to passenger congestion was not considered.

This study aims to develop an macroscopic model to estimate equilibrium distribution of passenger arrivals in a congested urban rail transit system. Passenger congestion influence and “knock-on delay” on the rail track are described by a train fundamental diagram (train-FD) sub-model first proposed by Seo et al.[30]. Due to its high tractability, the proposed model can serve as an effective tool to evaluate management strategies' influence on passengers' departure time choice behavior. More specifically, based on the travel cost assumption, the equilibrium travel time was first obtained. Then, given the scheduled timetable and equilibrium travel time, the dynamics of rail transit system was described. Next, equilibrium distribution of passenger arrivals was derived by substituting vehicle-based average density and flow of trains into the inverse function of train-FD. Finally, the characteristics of the equilibrium and applications of the proposed model are numerically described.

This paper is structured as follows. Section 2 briefly describes the scope of the proposed model, introduces the assumptions of passengers' travel cost and operation rules of rail transit. Section 3 formulates the equilibrium model, gives the solution method and existence conditions of the equilibrium. Section 4 describes the characteristics of the equilibrium in terms of rail transit performance and passenger arrival distribution. Section 5 applies the proposed equilibrium model to the management strategies including peak/off-peak timetable optimization and coarse pricing. Finally, conclusions and probable future works are briefly summarized in Section 6.

2. Model assumption

The rail transit system in this paper refers to a single urban railway line with multiple stations homogeneously located along the line. Meanwhile, it is assumed to be a strict First-In-First-Out (FIFO) system so that trains on the track cannot overtake each other. Besides, we make a major premise that travel time in rail transit system changes due to both passenger congestion and train congestion on the track. In addition, we assume that passengers can always board the next arriving train and will not change to other travel mode or give up their travel.

(1) Passenger travel cost

In this study, we assume that individuals mainly trade off the travel time cost (TTC) and schedule delay cost (SDC) to determine their departure time from home. The travel time from home to the nearest station for any individual is assumed to be constant so that it is equivalent to consider passengers' arrival time at the station. For the sake of clarity, "passenger departure" refers to the departure from the last station hereinafter. Therefore, the total travel cost of individual i when he or she departs the last station at time t is defined as:

$$TC(t, t_i^*) = \alpha(T(t) - T_0) + s(t, t_i^*) \qquad (1)$$

where t_i^* is the desired departure time from the last station for individual i. Here, t_i^* is also equivalent to the work start time for individual i if we take the time from the last station to workplace as a constant value. α ($\alpha > 0$) is value of time for congestion delay, $T(t)$ is the travel time of an individual when he or she departs the rail transit system at time t. For simplicity, $T(t)$ is the travel time of both trains and passengers if we set the boundary of the input-output system by the first and the last station for those commuting passengers. In this sense, this model can be applied to either one section or the whole rail transit line. But more practically, it is better to pick up the most congested section of a rail transit line and then consider the largest OD demand in this section. T_0 is the minimum travel time before the morning commute starts, and $s(t, t_i^*)$ is the schedule delay cost for individual i when he or she departs the system at time t. For the simplest case when all passengers have the same fixed desired departure time or work start time, t^*,

$t_i^* = t^*$ for any i.

Here, we employ the piece-wise linear schedule delay function widely used in most previous studies[18), 25), 29), 31), 32)]. Schedule delay $s(t, t_i^*)$ is defined as:

$$s(t, t_i^*) = \begin{cases} \beta(t_i^* - t), & \text{if } t \leq t_i^*, \\ \gamma(t - t_i^*), & \text{if } t > t_i^*, \end{cases} \tag{2}$$

where β and γ ($\gamma > \beta > 0$) are the value of time for early and late schedule delay, respectively. For simplicity, we assume that all the individuals have the same time cost α, β and γ.

(2) Rail transit operation

Denote the total length of the railway line as L and distance between adjacent stations as l. Trains' dwelling time t_b at each station is defined as:

$$t_b = t_{b0} + \mu a_p H, \tag{3}$$

where t_{b0} is the minimum dwelling time, μ is the rate of increase of dwelling time with boarding passengers' number on the platform, a_p is the passengers' arrival rate at platform to board the next arriving train, and H is the headway of two succeeding trains.

The cruising behaviour of trains are assumed to be described by the Newell's simplified car-following model[33)]. In this model, a vehicle either travels at its desired speed or follows the preceding vehicle while keeping the safety clearance. More specifically, position of train m at time t is described by

$$x_m(t) = \min\{x_m(t-\tau) + v_f\tau, x_{m-1}(t-\tau) - \delta\}, \tag{4}$$

where $m-1$ refers to the preceding train of train m, τ is the minimum time headway of successive trains, and δ is the minimum spacing. The first term represents the free-flow regime where the train cruises with its desired speed v_f. The second term indicates the congested regime where the train decreases its speed to maintain the minimum headway and spacing.

Now, the relation among train flow q ($q = 1/H$), train density k, and passenger arrival rate a_p in steady state, named as train fundamental diagram (train-FD), can be expressed as:

$$q = Q(k, a_p). \tag{5}$$

Based on the operation principles described in Eq. (3) and Eq. (4), the train-FD in Eq. (5) can be explicitly expressed as:

$$Q(k, a_\mathrm{p}) = \begin{cases} \dfrac{kl - \mu a_\mathrm{p}}{t_{\mathrm{b}0} + l/v_\mathrm{f}}, & \text{if } k < k^*(a_\mathrm{p}), \\ -\dfrac{\delta l}{(l-\delta)t_{\mathrm{b}0} + \tau l}(k - k^*(a_\mathrm{p})) + q^*(a_\mathrm{p}), & \text{if } k \geq k^*(a_\mathrm{p}), \end{cases} \tag{6}$$

where $q^*(a_p)$ and $k^*(a_p)$ are critical state train-flow and train-density, respectively, represented as:

$$q^*(a_p) = \frac{1 - \mu a_p}{t_{b0} + \delta/v_f + \tau}, \tag{7}$$

$$k^*(a_p) = \frac{(1-\mu a_p)(t_{b0} + l/v_f)}{(t_{b0} + \delta/v_f + \tau)l} + \frac{\mu a_p}{l}. \tag{8}$$

The derivation of Eq. (6) can be referred in the original paper by Seo et al.[30)].

3. User equilibrium

In this section, we first derive the passenger arrival rate under user equilibrium. Then, we show the solution method, and finally, the existence conditions of the equilibrium are briefly discussed.

(1) Derivation

The equilibrium state is reached when the total travel cost of any individual can not be reduced. Denote t_i as the departure time of individual i from rail transit under equilibrium, the following equation holds:

$$\frac{\partial TC(t_i, t_i^*)}{\partial t} = 0. \tag{9}$$

The derivative of travel time $TT(t)$ can therefore be obtained by substituting Eq. (1) and Eq. (2) into Eq. (9) as:

$$\frac{dT(t_i)}{dt} = \begin{cases} \beta/\alpha, & \text{if } t < t_i^*, \\ -\gamma/\alpha, & \text{if } t \geq t_i^*. \end{cases} \tag{10}$$

It can be observed that the values of travel time derivative do not depend on the desired departure time t_i^*. This is, however, not a general conclusion but a special result when piece-wise linear schedule delay function is employed. Eq. (10) implies that for passengers who depart the last station earlier than their desired departure time, the travel time they have experienced is increasing. On the contrary, for passengers who depart late, the travel time is decreasing. Since the derivatives of travel time in Eq. (10) are two constant values, travel time evolution under the equilibrium will be a piece-wise linear function with a single peak. If FIFO and First-In-First-Work (FIFW) hold, the travel time $T(t)$ maximizes at t_{max} when the schedule delay is zero. When all passengers have the same fixed desired departure time t^*, $t_{max} = t^*$. Finally, the travel time $T(t)$ under user equilibrium can be described as:

$$T(t) = \begin{cases} T_0 + \dfrac{\beta}{\alpha}(t - t_0), & \text{if } t_0 \leq t < t_{\max}, \\ T_0 + \dfrac{\beta}{\alpha}(t_{\max} - t_0) - \dfrac{\gamma}{\alpha}(t - t_{\max}), & \text{if } t_{\max} \leq t \leq t_{\mathrm{ed}}, \end{cases} \tag{11}$$

where t_0 and t_{ed} represent the start and end time of the equilibrium, respectively. Meanwhile, t_0 and t_{ed} are also the time when commuters start and finish departing the rail transit system, respectively.

In order to derive the equilibrium distribution of passenger arrivals, the dynamic traffic states of the rail transit system should be described. We first consider the rail transit system as a input-output system that can be approximately described by a fluid model. Denote $a(t)$ and $d(t)$ as the inflow and outflow of trains. Accordingly, $A(t)$ and $D(t)$ are the cumulative numbers of $a(t)$ and $a(t)$, respectively. According to FIFO, the cumulative departures of trains at t should be equal to the cumulative arrivals of trains at $t - T(t)$, which can be written like:

$$D(t) = A(t - T(t)). \tag{12}$$

Differentiate the both sides of this equation we can get

$$d(t) = a(t - T(t))\left(1 - \frac{dT(t)}{dt}\right). \tag{13}$$

Then, it is natural to assume that average density and flow of trains follow the steady state train-FD introduced in Section 2(2). Thus, the equilibrium passenger arrival rate can be calculated by the inverse function of train-FD as:

$$a_p = Q^{-1}\left(\bar{k}(t), \bar{q}(t)\right). \tag{14}$$

The ground of this assumption is that by aggregating the density and flow, their sudden change with time can be flattened, and the traffic states represented by the averaged variables can approximately be regarded as the steady state.

Here, we adopted the vehicle-based harmonic mean to be applied in Eq. (14) since it is simple and tractable. More specifically, the average flow $q(n)$ experienced by train n (or equivalently average headway $H(n)$) can be written as:

$$q(n) = \frac{1}{H(n)} = \left(\frac{H_a(n)+H_d(n)}{2}\right)^{-1}. \tag{15}$$

Where $H_a(n)$ and $H_d(n)$ are the time headways of train $n-1$ and train n (simply referred to as the headway of train n hereinafter) at the first arrival station and the last departure station, respectively. According to the definition, $H_a(n)$ and $H_d(n)$ can be further expressed as:

$$H_a(n) = \frac{d(t-TT(t))}{dn} = \frac{1}{a(t-T(t))}, \tag{16a}$$

$$H_d(n) = \frac{dt}{dn} = \frac{1}{d(t)}. \tag{16b}$$

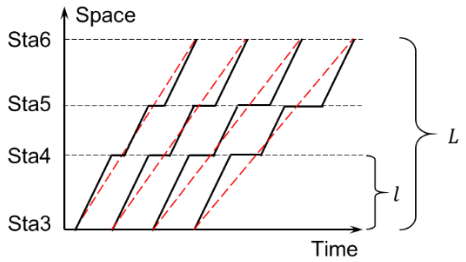
Similarly, the average density $k(n)$ (or equivalently average spacing $s(n)$) can be written as:

$$k(n) = \frac{1}{s(n)} = \left(\frac{s_a(n)+s_d(n)}{2}\right)^{-1}, \tag{17}$$

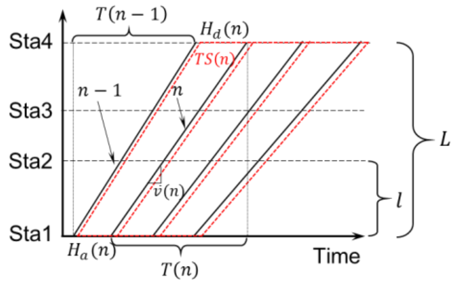
$$s_a(n) = \bar{v}(n)H_a(n) = \frac{L}{T(n)}H_a(n), \tag{18a}$$

$$s_d(n) = \bar{v}(n)H_d(n) = \frac{L}{T(n)}H_d(n). \tag{18b}$$

Where $\bar{v}(n)$ is the average travelling speed of train n and $T(n)$ is the travel time of train n from the first station to the last station. For a more intuitive understanding, we specify the definitions of these variables by an illustration in **Fig. 1**.



(a) Trajectories of trains during morning commute



(b) Simplified trajectories of trains

Fig. 1 An example of trajectories of trains and definition of variables.

Fig. 1(a) gives an example of trajectories of trains operating on a railway line with four stations from "Sta3" to "Sta6". When the morning commute starts, passenger arrival rate increases with time so that the dwelling time of trains is gradually extended at each station as can be understood from **Fig. 1**(a). To describe the average density and flow, the trajectories are simplified by connecting the start and end point on the time-space diagram. Then, the whole trip of a train n can be macroscopically described by its average travelling speed $\bar{v}(n)$ or travel time $T(n)$ as shown in **Fig. 1**(b). In addition, if the time-space diagram is delimited as the trapezium in red dotted line like $TS(n)$ in **Fig. 1**(b), you may find that the definitions of $q(n)$ and $k(n)$ in Eq. (15) and Eq. (17) are consistent with Edie's definition in a time-space diagram[34].

In a single route FIFO rail transit system, another relation between headways and travel time for two succeeding trains $n-1$ and n should hold as:

$$H_d(n) - H_a(n) = T(n) - T(n-1). \tag{19}$$

Now, when the inflow $a(t)$ (or equivalently $H_a(n)$ for all $n \in N$) is given and equilibrium travel time $T(t)$ is obtained from Eq. (11), outflow $d(t)$ and $H_d(n)$ can be obtained by Eq. (13) and Eq. (16b). Then, vehicle-based travel time $T(n)$ can be calculated by Eq. (19) with an initial $T(1)$. Next, the average density $k(n)$ and flow $q(n)$ for all $n \in N$ can be calculated by Eq. (17) and Eq. (15), respectively. Finally, passenger arrival rate under equilibrium can be derived by substituting the $q(n)$ and $k(n)$ into Eq. (14) as:

$$a_p(n) = Q^{-1}\big(k(n), q(n)\big), \tag{20}$$

where $a_p(n)$ can be interpreted as the time-averaged passenger arrival rate during the trip of train n.

For a specific morning commute situation, boundary conditions have to be introduced to get feasible results. For simplicity, we only consider the case when all passengers have the same desired departure time (or work start time). The inverse problem in Eq. (20) can be solved if the start and end time (i.e., t_0 and t_{ed}) of the morning commute are determined. Therefore, two boundary conditions should be introduced. The first one is that the travel time when commuters finish departing the rail transit system should be the same with the initial travel time before morning commute starts, as described by Eq. (21). The second one is that there is a total travel demand denoted by N_p for commuters as described by Eq. (22).

$$T(t_{ed}) = T_0 = \frac{L}{l}\left(t_{b0} + l/v_f\right), \tag{21}$$

$$\int_1^N a_p(n)\, dn = N_p. \tag{22}$$

In addition, $a_p(n)$ calculated from Eq. (20) should be adjusted so that commuters only arrive during $[t_0, t_{ea}]$ (i.e., Eq. (23)), where t_{ea} is the time when final commuters arrive the rail transit system as explained in Eq. (24).

$$\begin{cases} a_p(n) \geq 0, & \text{if } t_0 \leq t(n) \leq t_{ea} \\ a_p(n) = 0, & \text{otherwise} \end{cases}, \tag{23}$$

$$t_{ea} = t_{ed} - T(t_{ed}) = t_{ed} - T_0, \tag{24}$$

(2) Solution method

The calculation process to solve the problem is shown in Algorithm 1, where Δ is the step size of time, and N_ε is the tolerance limit of passenger number.

Algorithm 1 Solution

Input: Operational parameters l, L, t_{b0}, μ, v_f, δ, τ; cost function $TC(t, t^*)$, train inflow $a(t)$, total travel demand N_p;

Output: Train flow $q(n)$, train density $k(n)$, and passenger arrival rate $a_p(n)$;

1: Set an initial t_0;
2: Calculate t_{ed} and t_{ea} by Eq. (21) and Eq. (24);
3: Calculate $T(t)$ and $d(t)$ by Eq. (11) and Eq. (13);
4: Calculate $a_p(n)$ by Eq. (20) together with Eq (15~19);
5: Adjust $a_p(n)$ according to Eq. (23);
6: Calculate $LHS - RHS$ of Eq. (22), denoted as $error$;
7: **if** $error < -N_\epsilon$, **then**
8: $\quad t_0 = t_0 - \Delta$, repeat Step 2 to 6;
9: **else if** $error > N_\epsilon$, **then**
10: $\quad t_0 = t_0 + \Delta$, repeat Step 2 to 6;
11: **else**
12: $\quad$ Calculation converges, t_0, t_{ea} and t_{ed} determined;
13: **end if**
14: Outputs are derived from Step 4 when the calculation converges.

---

(2) Existence conditions of equilibrium

When the travel time under the equilibrium and train-FD are jointly employed, several constraints have to be imposed to ensure the calculation results are physically feasible. In fact, train-FD is a bounded set in which all feasible traffic states (i.e., sets of $\{k, a_p, q\}$ that conform to Eq. (6)) are included. In other words, any traffic state outside of train-FD cannot be reached due to the physical limits of a rail transit system.

The first constraint is that the outflow of trains should always be positive, which means:

$$d(t) = a\big(t - T(t)\big)\left(1 - \frac{dT(t)}{dt}\right) > 0, \forall t. \tag{25}$$

Train inflow $a(t)$, in this study, is treated as a positive given input so that $1 - dT(t)/dt > 0$ should hold all the time. Substitute Eq. (10) into this formula we can get

$$\alpha > \beta. \tag{26}$$

This constraint is consistent with most equilibrium models for road traffic (e.g., Hendrickson and Kocur[18]; Arnott et al.[35]).

The second constraint is that passenger arrival rate $a_p(n)$ calculated from Eq. (20) should not be negative as written in Eq. (27).

$$a_p(n) \geq 0,\ \forall n. \tag{27}$$

In fact, this constraint is equivalent to that the set of $(k(n), q(n))$ calculated from equilibrium travel time should not exceed the boundary of train-FD. However, the evolution of $(k(n), q(n))$ depends on not only the constant operational parameters, but also the settings of train inflow and total travel demand. To obtain an explicit expression of Eq. (29), we consider a simplified case when inflow is constant (i.e., $a(t) \equiv a_c$). From the discussion in Section 4 it can be understood that as long as the upper right corner of $(k(n), q(n))$ loop does not exceed the congested regime boundary of train-FD, the whole loop of $(k(n), q(n))$ will lie inside the train-FD. According to the equilibrium condition, the upper right corner of $(k(n), q(n))$ loop refers to the state when the travel time is just maximized and starts to decrease. Therefore, the right corner $k_{rc}(n)$ and $q_{rc}(n)$ can be written as:

$$k_{rc}(n) = \frac{T_{max}}{\left(\frac{1}{a_c}+\frac{1}{a_c(1+\gamma/\alpha)}\right)\cdot L/2}, \tag{28}$$

$$q_{rc}(n) = \frac{1}{\left(\frac{1}{a_c}+\frac{1}{a_c(1+\gamma/\alpha)}\right)}, \tag{29}$$

Then, the constraint that $k_{rc}(n)$ and $q_{rc}(n)$ does not exceed congested regime boundary of train-FD can be derived by substituting $k_{rc}(n)$ and $q_{rc}(n)$ into lower branch of Eq. (6) as:

$$\frac{\alpha+\gamma}{\alpha+\gamma/2} a_c \left[\frac{T_{max}}{L} + \frac{(l-\delta)t_{b0}+\tau l}{\delta l}\right] \leq \frac{1}{\delta}, \tag{30}$$

Where T_{max} is an indicator of total travel demand since t_0 is negatively correlated with total travel demand as can be understood from Eq. (11).

Finally, Eq. (26), Eq. (27) can be considered as the existence conditions of the equilibrium while Eq. (30) gives an explicit expression of Eq. (27) for a simplified case when train inflow is constant.

4. Characteristics of the equilibrium

In this section, the characteristics of the equilibrium by the proposed model are discussed by a series of numerical examples. The parameter settings for the numerical examples are first introduced. Then, dynamics of rail transit and passenger arrival distribution under equilibrium for both high and low demand conditions are interpreted.

The parameter settings used in the following numerical examples are listed in **Table 1**. For simplicity, the train inflow $a(t)$ is set as constant.

Table 1. Parameter settings for numerical example.

Parameter	Value	Parameter	Value
l	1.2 km	α	20 $/$h$
L	18 km	β	8 $/$h$
v_f	40 km/h	γ	25 $/$h$
t_{b0}	20 sec	t^*	240 min
μ	0.1 sec/pax	$a(t)$	12 tr/h
δ	0.4 km	N_p	30000 pax
τ	1.0 min	N_ϵ	100 pax
Δ	1.0 min		

Fig. 2 shows the costs under the equilibrium (i.e., during $[t_0, t_{ed}]$). Total travel cost $TC(t,t^*)$ under equilibrium is minimized and constant so that no one has the incentive to change his time choice behaviour. To achieve the constant total travel cost, travel time cost should first increase until t^* and then decreases to t_{ed}.

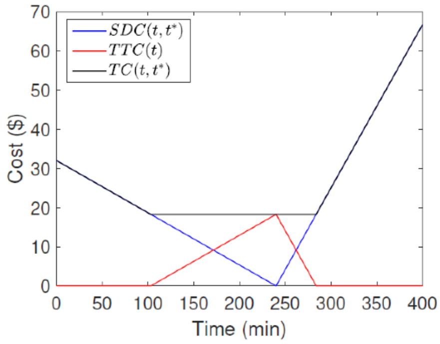


Fig. 2 Travel costs under equilibrium.

Fig. 3 depicts the cumulative curves of trains. It can be seen that $D(t)$ will first diverge from $A(t)$ from t_0 to t^* and then approach to $A(t)$ from $t^*$to t_{ed}. In this way, the travel time is under user user equilibrium as described in Eq. (11).

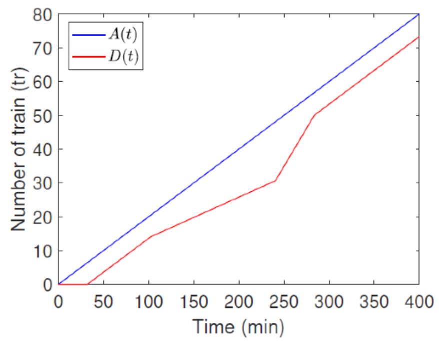


Fig. 3 Cumulative number of trains.

Fig. 4 shows the cumulative arrival and departure of passengers, the difference of two adjacent circles on the curve indicates the number of passengers carried by the preceding train. It can also be observed that riding passenger numbers on trains before and after t^* significantly vary under the constant train inflow setting.

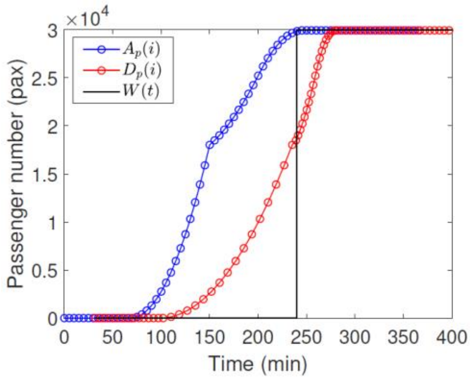


Fig. 4 Cumulative number of passengers ($N_p = 30000$).

Passenger arrival rate $a_p(i)$ and departure rate $a_p(i)$ are depicted in **Fig. 5**. It can be observed that $a_p(i)$ has two peaks in this example, the large one corresponds to the arrival of commuters who experienced the longest travel time but departed the system just before t^* while the small one corresponds to the arrival of commuters who experienced shorter travel time at the cost of departing the system after t^*. Since $a_p(i)$ is derived from the inverse function of train-FD, how $k(i)$ and $q(i)$ evolve on train-FD is also drawn in **Fig. 6**. It can be seen that the evolution of $(k(i), q(i))$ during the morning commute starts from the left boundary of train-FD and formulates a counter-clockwise closed loop within the train-FD. Divided by $a(t) = 12\ tr/h$, the lower half of the loop represents the average density and flow experienced by trains departing the system from t_0 to t^*, while the upper half represents the dynamics from t^* to t_{ed}. The maximum of a_p is reached at the lower right corner of the loop. In fact, if train inflow $a(t)$ is constant and late departure in the schedule delay function (i.e., the condition $t > t_i^*$ in Eq. (2) is valid) is permitted, passenger arrival rate with two-peaks appear when high travel demand forces $(k(i), q(i))$ to enter the congested regime of train-FD.

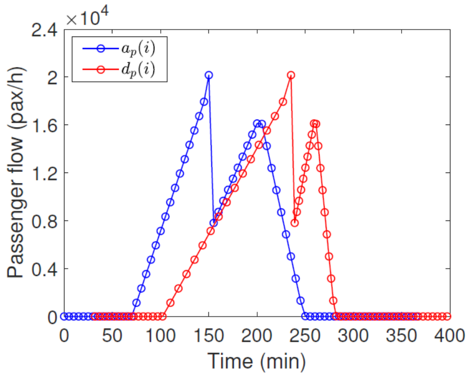


Fig. 5 Passenger flow ($N_p = 30000$)

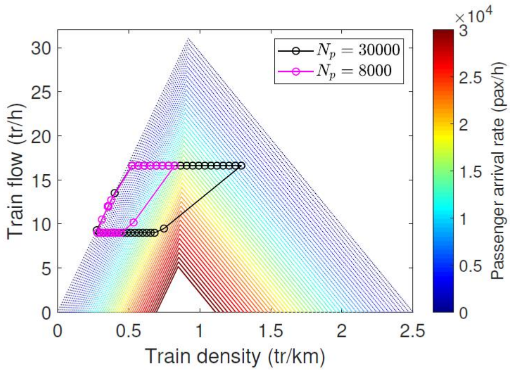


Fig. 6 Dynamics of density and flow on train-FD

On one hand, if late departure is not permitted (i.e., only the condition $t \leq t_i^*$ in Eq. (2) is valid), the travel time under the equilibrium will continue increasing until t^* so that $k(i)$ calculated from Eq. (17) also keeps increasing. As a result, $(k(i), q(i))$ goes deeper into the train-FD from the left side so that a_p increases in the free-flow regime and decreases in the congested regime. Then, after t^*, no commuter departs from the system (i.e., $a_p = 0$) so that $(k(i), q(i))$ suddenly jumps up to the upper boundary of train-FD and finally returns to the initial point along the boundary of the train-FD. In this way, a_p will only have one peak.

On the other hand, if the late departure is permitted but the total travel demand is rather low, a_p may also have only one peak. **Fig. 7** and **Fig. 8** give an example when a_p has one peak under the travel demand $N_p = 8000$. The comparison of $(k(i), q(i))$ evolution between high and low demand is also shown in **Fig. 6**. It can be understood that if $(k(i), q(i))$ after t^* (upper right of the loop) does not enter the congested regime of train-FD, there will be only one peak for a_p. In other words, commuters who departed the system after t^* are represented in the decline side of a_p if total travel demand is low. On the contrary, when the demand is high, arrivals of early and late commuters are separated on the two peaks. This is a rather new and interesting characteristic of equilibrium distribution of passenger arrivals for a rail transit system.

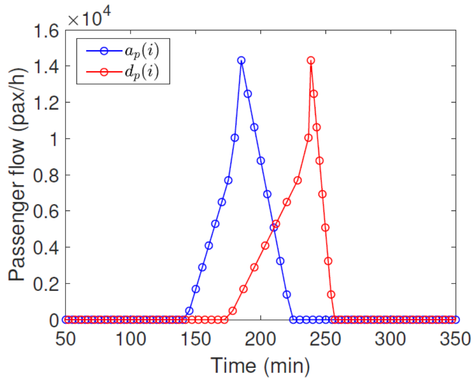


Fig. 7 Passenger flow ($N_p = 8000$)

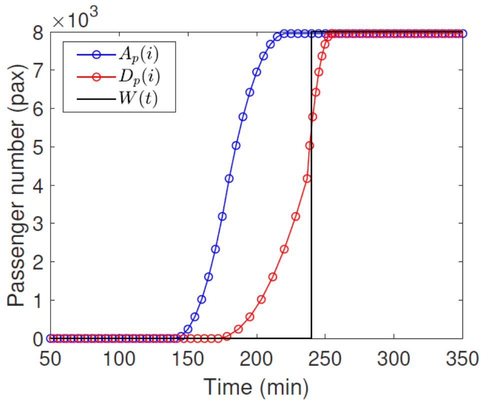


Fig. 8 Cumulative number of passengers ($N_p = 8000$)

5. Model applications

By applying the proposed equilibrium model, this section aims to quantitatively analyze the policy implications of timetable optimization and coarse pricing, which are expected to derive insights into more efficient operation of railway system and flattening the surging passenger demand during rush hours. More specifically, the conventional peak/off-peak timetable is numerically optimized under user equilibrium in the first subsection. A single-step coarse pricing scheme (including surcharge, reward and their combinations) is numerically evaluated based on indicators such as social cost, user cost and in-vehicle crowding.

(1) Peak/off-peak timetable optimization

In this section, the optimization problem of a conventional peak/off-peak timetable without pricing is discussed. Peak/off-peak timetable here refers to a simplified timetable pattern that

- Only two types of dispatch frequencies (or inflow, referred to as $a_1 \geq a_2$) are employed;
- The high dispatch frequency lasts for an uninterrupted period enclosed by the low dispatch frequency.

Since this timetable pattern is widely employed in practice, it is desirable to explore the optimal setting of peak/off-peak timetable by the proposed model. This issue can be generalized as a bi-level optimization problem described as:

$$\min_{a_1,a_2\in\Omega} C^e = \min_{a_1,a_2\in\Omega} f\left(a_1, a_2 | N_p\right), \qquad (31)$$

where C^e is the equilibrium cost, which is the same for all commuters under user equilibrium. Ω should be an enclosed area for two inflows. Here, for simplicity, we consider an ideal case where trains that carry commuters and arrive the destination before t^* are dispatched with inflow a_1, and other trains are dispatched with inflow a_2, as shown in **Fig. 9**.

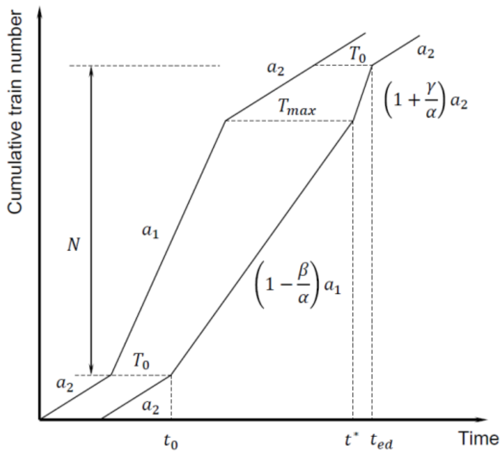


Fig. 9 An ideal case of train cumulative curve for peak/off-peak timetable pattern.

This ideal case is also reasonable because under the equilibrium, train ouflow before t^* would be smaller than the inflow due to increasing passenger arrivals. On the contrary, outflow after t^* would be larger than the inflow due to decreasing passenger arrivals. Therefore, to allocate high inflow a_1 to trains departing before t^* would help maintain a higher level of average flow of railway system, while to allocate low inflow a_2 after t^* help avoid bunching phenomenon due to sudden decrease of passenger arrivals. Under this ideal case, the available area for a_1 and a_2, Ω, can be expressed by the following four constraints:

$$s.t.\ 0 < a_2 \leq a_1 \leq 1/\tau, \qquad (32)$$

$$s.t.\ \omega a_1 + (1-\omega)a_2 \leq a_0, \qquad (33)$$

$$s.t.\ a_1 \leq \frac{\alpha-\beta/2}{\alpha-\beta}\left(\frac{\delta}{L}T_{\max} + \left(1-\frac{\delta}{l}\right)t_{b0} + \tau\right)^{-1} = a_1^*. \qquad (34)$$

$$s.t.\ a_2 \leq \frac{\alpha+\gamma/2}{\alpha+\gamma}\left(\frac{\delta}{L}T_{\max} + \left(1-\frac{\delta}{l}\right)t_{b0} + \tau\right)^{-1} = a_2^*. \qquad (35)$$

The first constraint clarifies that train inflow can not be larger than the reciprocal of minimum time headway τ. The second constraint limits the achievable train inflow (due to available train fleet) as a_0. The third and fourth constraints comes from the feasible density and flow within train-FD, which is related to demand and has been explained in Eq. (30). In fact, ω can be explicitly derived from the duration ratio for a_1 and a_2 as:

$$\omega = \frac{\gamma(\alpha-\beta)}{\beta(\alpha+\gamma)+\gamma(\alpha-\beta)}, \qquad (36)$$

$$1-\omega = \frac{\beta(\alpha+\gamma)}{\beta(\alpha+\gamma)+\gamma(\alpha-\beta)}. \qquad (37)$$

In fact, to explicitly express equilibrium cost C^e as a function of a_1, a_2 and N_p is tedious since it relates to train-FD and dynamic model. Here, a numerical solution is given to the optimization problem. The idea is to enumerate all feasible combinations of a_1, a_2 that conforms to the first constraint. Then, run similar algorithm introduced in Algorithm 1. It can be observed from **Fig. 9** that when t_0 is given, temporal setting of a_1 and a_2 can be determined, which means $a(t)$ is obtained. Thus, the following steps are almost the same with Algorithm 1, only to add a check of third and fourth constraints as written in Eq. (36) and Eq. (37). The core part of the algorithm is a three-layer iteration process. The outer layer iterates $a_1 < 1/\tau$, the medium layer iterates $a_2 \leq a_1$, the inner layer iterates t_0 to minimize N_p error. Under some extreme combinations of a_1 and a_2, the third and fourth constraints cannot be met or the N_p error is too large, which is considered to be not acceptable case.

Here, an numerical example of optimization which still uses the parameter settings in **Table 1** is given. The iteration sets the step length of inflow as 0.1 tr/h, and in a range of $[5, 30]$ tr/h. N_p error tolerance is 500. **Fig. 10** shows the calculation result, with a_1 and a_2 as horizontal and vertical axes color mapped by equilibrium cost C^e. It can be seen that equilibrium cost C^e minimized in an enclosed area of $a_1$around $[16,20]$ tr/h and a_2 around $[9,11]$ tr/h. This also confirms that to adopt an high dispatch frequency enclosed by low dispatche frequency is reasonable from the perspective of reducing equilibrium travel cost.

In addition, the constant train inflow situation discussed in Section 4 is represented by the red point in **Fig. 10**. It can be calculated that compared to constant case, the optimized peak/off-peak timetable could reduce the equilibrium cost up to around 18%, which is rather significant. Besides, the second constraint can be represented by the left side of the black line in **Fig. 10** depending on the value of a_0.

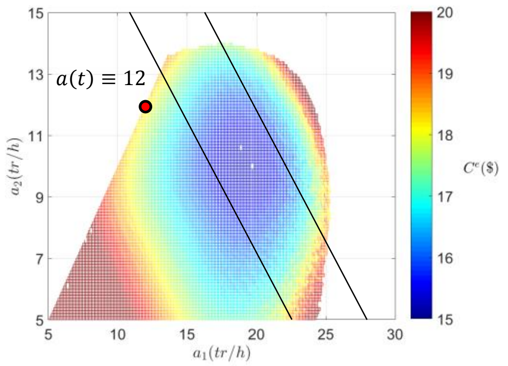


Fig. 10 An example of optimization result for peak/off-peak timetable pattern.

(2) Coarse pricing

In this section, a practical pricing scheme: single-step coarse pricing is numerically evaluated using the proposed model. Since the first-best (or continuous) pricing involves practical difficulties in collecting time-dependent fare or giving time-dependent reward, an approximation, step-tolling, has been widely discussed for road traffic context[35)–37)]. With regard to rail transit, studies on fare scheme design and optimization are still at the beginning stage compared to road traffic[29), 38), 39)]. When a single-step coarse pricing scheme is introduced, user cost (UC) can be written as the following equation (for simplicity, only consider the case when $t_i^* = t^*$ for any i):

$$UC(t,t^*) = \alpha\,(T(t) - T_0) + s(t,t^*) \pm p,\ \text{if}\ t \in [t^+, t^-], \tag{38}$$

where $+p > 0$ refers to the single-step surcharge, $-p < 0$ refers to the single-step reward. t^+ and t^- are the start and end time of the pricing period, respectively. In order to keep UC unchanged during$[t^+, t^-]$, sudden change of TTC will inevitably appear at t^+ and t^-. An illustration of costs when coarse pricing is applied is shown in **Fig. 11**.

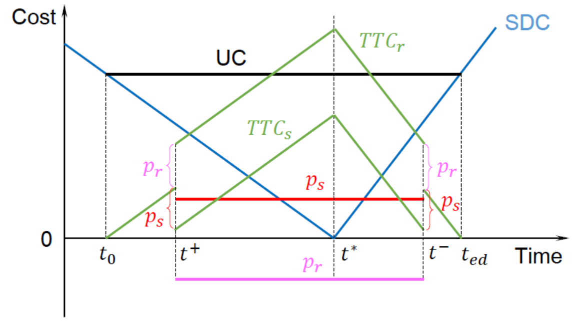


Fig. 11 Travel costs when coarse pricing applied.

Further, since TTC is proportional to travel time $T(t)$, the change of travel time at time boundary of pricing scheme $\Delta T = p/\alpha$ has to be appropriately interpreted. For road traffic, Arnott et al.[35)] proposed the concept of mass departures, Laih[36)] assumed the existence of additional lanes for drivers to wait. Lindsey et al.[37)] considered a more practical braking model to describe this phenomenon. For rail transit, mass departures are physically not possible since trains are cruising on the same track. Also, multiple trains ($\Delta n \geq 3$) departing with minimum headway to realize ΔT (i.e., $\Delta n(H_d - \tau) = p/\alpha$) will break the equilibrium requirement since at least two trains would appear at one side of time boundary with minimum headway τ, which is different from H_d. Note, the coarse pricing is not valid for the situation when trains are departing with minimum headway. The only feasible situation is that the change of travel time due to pricing is carried out by the two trains departing just before and after the time boundary. An example of surcharge is shown in **Fig.12**. It can observed that the headway of train $m + 1$ and $n + 1$ departing just after t^+ and t^- decreases/increases p/α. In this way, the sudden change of TTC is realized without breaking the equilibrium. Meanwhile, a constraint of p/α exists due to the FIFO and minimum headway as:

$$\frac{p}{\alpha} \leq H_d - \tau, \tag{39}$$

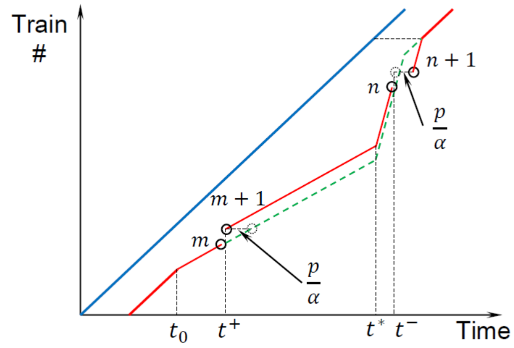


Fig. 12 Cumulative curves when surcharge imposed during $[t^+, t^-]$.

Then, when an appropriate pricing value p is set, solution to the equilibrium model is basically the same with original problem, but only modifies the travel time with $\Delta T = p/\alpha$ and corresponding $k(n)$ for trains departing between $[t^+, t^-]$. The effectiveness of coarse pricing are evaluated by the following three indicators in this study:

- Social cost (SC) change (%) compared to no pricing situation;
- User cost (UC) change (%) compared to no pricing situation;
- Standard deviation (STDEV) of riding passenger number change (%) compared to no pricing situation.

SC in this study refers to the sum of TTC and SDC. First, SC change evaluates whether the social welfare is improved or not since the monetary circulation between passengers and the operator is neutral from a social perspective. Second, UC change assess whether passengers benefit or not when the pricing scheme is implemented. Third, the change of riding passenger number n_p STDEV is employed to check whether the in-vehicle crowding is relieved or not because STDEV of n_p reflects the evenness of passenger distribution among dispatched trains. The decrease of this STDEV indicates the difference of n_p between congested and uncongested trains is narrowed.

Since the effect of pricing value is obvious and limited by Eq. (39), it is more interesting to examine how the temporal settings affect the effectiveness of the coarse pricing. The duration (i.e., $t^- - t^+$) and end time of the coarse pricing are picked here as the indicators of temporal settings. **Fig. 13** and **Fig. 14** show a series of numerical experiments of SC and UC change for the surcharge case $p = 2\$$. The train inflow in these scenarios takes the constant value $a(t) = 12\ tr/h$. The color represents the percentage of cost change as illustrated by the legend. On one hand, it can be observed from **Fig. 13** that SC change strongly depends on the end time of surcharge. When the surcharge covers the both side of t^* and ends at an appropriate time after t^* (in this example 20 min), its effect is maximized. This is reasonable because when surcharge ends before t^*, the most congested trains carrying passengers who want to be on time are not surcharged, thus the reduction of TTC is weakened. On the other hand, the user cost in **Fig. 14** basically increases as long as the surcharge is imposed. The longer the duration, the earlier or later it ends, the higher user cost would be.

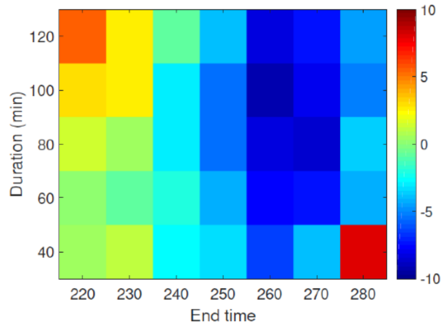


Fig. 13 Social cost change (%) for surcharge case

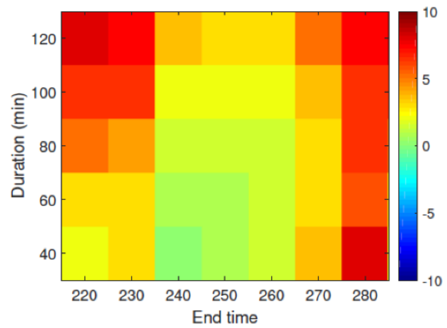


Fig. 14 User cost change (%) for surcharge case

To further interpret the interior reason of UC and SC change, UC is decomposed to the sum of TTC, SDC and total pricing cost (TP). **Fig. 15** and **Fig. 16** show an example of each cost component change under a given duration (i.e., $t^- - t^+ = 100\ min$) or under a given end time (i.e., $t^- - t^* = 20\ min$). On one hand, it can be observed from **Fig. 15** that when the end time of 100 min surcharge moves from the left side to the right side of t^*, TP significantly increases while TTC first decreases until 260 min (i.e., 20 min after t^*) and then inversely increases. The increase of TP means more passengers are surcharged. The change of TTC indicates that there may exist an optimal end time of the surcharge, if it ends further later, the surcharge will cover all late departed commuters and forces some of them to use earlier trains including those that have already been heavily congested. As a result, riding passenger number difference is not efficiently reduced and TTC inversely increases. SDC in this comparison slightly increases or decreases and does not show a consistent result. On the other hand, when the end time of the surcharge is fixed to cover the both sides of t^* and the duration increases as shown in **Fig. 16**, it can be observed that the TTC slightly declines and then rises with duration extension while SDC almost remains unchanged. This indicates that when the surcharge coves an appropriate period of both sides of t^*, it is not efficient to excessively extend the duration of surcharge (in this example, a duration around 100 min is optimal).

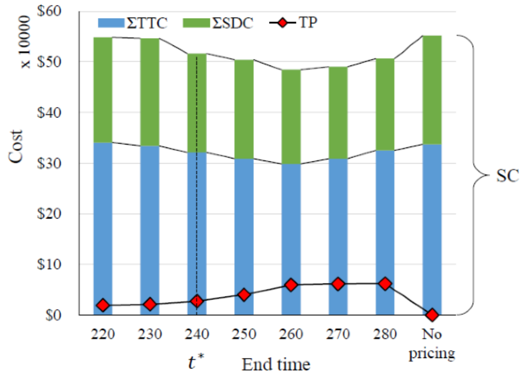


Fig. 15 Cost breakdown when $t^- - t^+ = 100\ min$ for surcharge case.

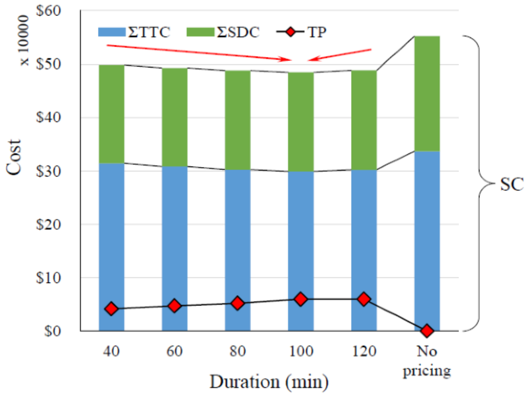


Fig. 16 Cost breakdown when $t^- - t^* = 20\ min$ for surcharge case.

Similarly, the sensitivity of cost change is also conducted for the reward case as shown in **Fig. 17** and **Fig. 18** (Reward $p = 2\$$ only given before t^* and train inflow still takes the constant value $a(t) = 12\ tr/h$). On one hand, it can be understood from **Fig. 17** that SC change varies in a relatively small range compared to surcharge depending on both end time and duration, and there is an optimal end time and duration around (170 min, 90 min). When the end time of reward approaches t^*, SC inversely increase. This is because giving reward to passengers who should be surcharged according to first-best pricing would encourage passengers to utilize congested trains, as a result, TTC and SC increase. On the other hand, UC generally decreases as long as the reward is given as can be confirmed from **Fig. 18**. The longer the duration, the earlier it ends, the lower user cost would be. However, there also exists a limit for the reduction of UC around the left top of **Fig. 18** since giving reward before the start of commute does not have effect.

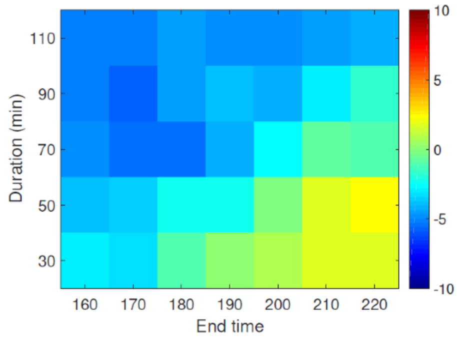


Fig. 17 Social cost change (%) for reward case.

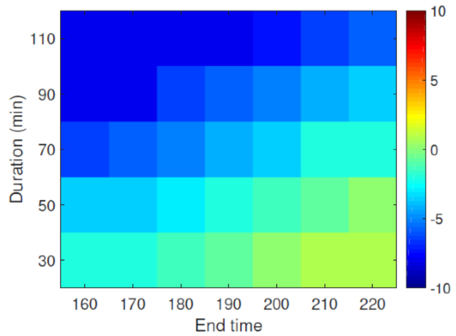


Fig. 18 User cost change (%) for reward case.

An example of cost breakdown for reward case is shown in **Fig. 19** and **Fig. 20** to further understand the reward effect on different components of the cost. On one hand, when the duration of the reward is fixed, it can be seen from **Fig. 19** that TTC significantly decreases when the reward ends earlier before t^*. This is because the reward encourages some of the passengers to use the earlier trains, as a result, the difference of riding passenger number between congested and uncongested trains narrows, thus TTC declines. In addition, it can also be observed that when the reward ends earlier, SDC increases. This is reasonable because when more passengers are encouraged to ride on earlier trains, their schedule delay undoubtedly raise. Therefore, when the increase of SDC just exceeds the decrease of TTC, SC is minimized (in this example around 170 min). On the other hand, when the end time of reward is fixed as 70 min before t^* in **Fig. 20**, TTC decreases with the extension of the reward period. The reason is the same as explained before. However, the minimum of SDC appears at 70 min, which is also the minimum of SC. This result indicates that the use of reward is not the longer the better. If the reward period is too long, the increase of SDC may exceed the decrease of TTC. Therefore, an appropriate period of reward sufficiently earlier before t^* is the most desirable situation.

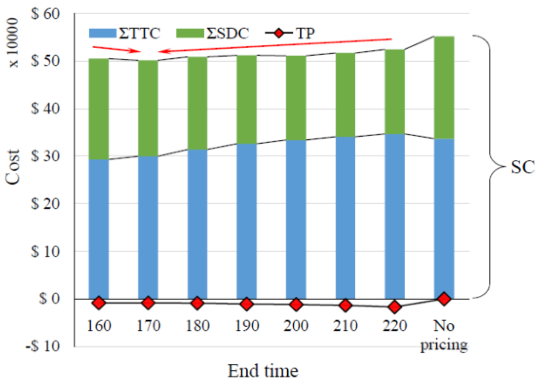


Fig. 19 Cost breakdown when $t^- - t^+ = 90\ min$ for reward case.

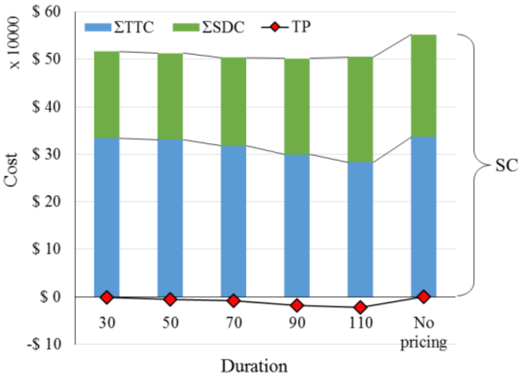


Fig. 20 Cost breakdown when $t^- - t^* = -70\ min$ for reward case.

In addition, change of n_p STDEV for both surcharge and reward cases are shown in **Fig. 21** and **Fig. 22**. It can be observed from **Fig. 21** that when surcharged is imposed, the change of n_p STDEV behaves similar as that of SC in **Fig. 13**. The decrease of n_p STDEV indicates that the in-vehicle crowding is more evenly distributed. Differently, n_p STDEV change for reward case seems sensitive to both the pricing duration and end time as can be understood from **Fig. 22**. n_p STDEV even significantly increases when the reward lasts 30 min and ends 20 min before t^*. To conclude, the change of n_p STDEV is basically consistent with SC change, which implies that to reduce the SC basically will also alleviate in-vehicle crowding.

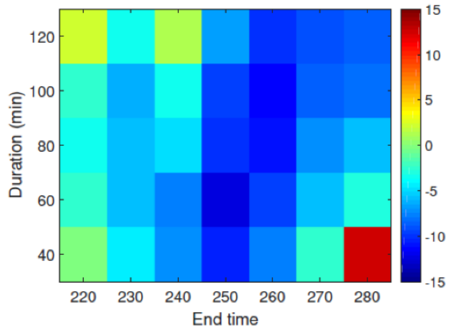


Fig. 21 STDEV of n_p change for surcharge case (%).

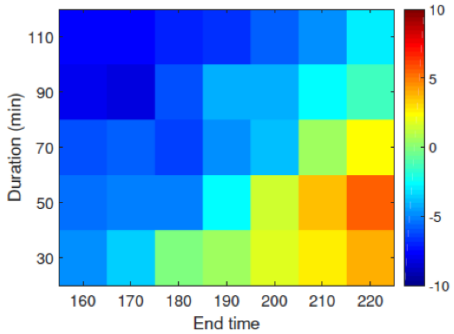


Fig. 22 STDEV of n_p change for reward case (%).

6. Conclusions and future work

This paper proposed an macroscopic model to estimate equilibrium distribution of passenger arrivals for the morning commute problem in a congested urban rail transit system. Compared to the existing equilibrium models, this model is the first to incorporate the interaction between passenger congestion and operational performance of trains by employing a train-FD sub-model.

With the proposed model, we numerically analysed the equilibrium distribution of passenger arrivals and found that total travel demand and train timetable all had significant influence on the characteristics of the distribution. More specifically, if all passengers have to depart the rail transit system no later than their desired departure time, there will be only one peak for the equilibrium distribution of passenger arrivals, otherwise, there may exist two peaks depending on the travel demand. When the travel demand is low so that train density and flow do not enter the congested regime of train-FD, the equilibrium distribution will also be unimodal.

By applying the proposed model, a conventional peak/off-peak timetable pattern is numerically optimized under user equilibrium. It is found that compared to the constant dispatch frequency, the combination use of high and low dipatch frequencies could significantly reduce the equilibrium cost for commuters. The implications from the numerical experiments of a single-step coarse pricing can be concluded as:

- By implementing one-step coarse pricing without breaking the user equilibrium, social cost can be acceptably reduced.
- The surcharge period should cover both sides of t^*, while its duration is not necessary to be excessively long.
- The reward period should end before t^* with an appropriate time interval, while its duration has an optimal length from the perspective of social cost.

There are still many issues that can be improved to further strengthen the reliability of the findings above and make the insights more applicable to practice. Here, three directions of future work can be considered. First, improvement and further investigation of the train-FD should be conducted. Traffic flow model used in this study assumes a homogeneous railway system, which can be improved to be more realistic by developing a train-FD model also applicable for heterogeneous railway system. Second, the equilibrium can be extended from two aspects. One is to consider the heterogeneity of commuting passengers which means the passenger's value of time should be various and group-dependent. Another aspect is to consider including the in-vehicle crowding cost into the cost function. Third, two probable issues can be considered to improve the practicality of the proposed model. One is to expand the single railway line system to simple network since passengers' transfer at major stations plays a non-negligible role in morning commute. Another one is to conduct case studies on timetable optimization and pricing schemes with the proposed model.

REFERENCES

1) Vuchic, V. R. (2017). Urban transit: operations, planning, and economics. John Wiley & Sons.
2) Xu, X., Li, H., Liu, J., Ran, B., and Qin, L. (2019). Passenger flow control with multistation coordination in subway networks: algorithm development and real-world case study. *Transportmetrica B: Transport Dynamics*, 7(1):446–472.
3) Kariyazaki, K., Hibino, N., and Morichi, S. (2015). Simulation analysis of train operation to recover knock-on delay under high-frequency intervals. *Case Studies on Transport Policy*, 3(1):92–98.
4) Carey, M. and Kwiecinski, A. (1994). Stochastic approximation to the effects of headways on knock-on delays of trains. *Transportation Research Part B: Methodological*, 28(4):251–267.
5) Tirachini, A., Hensher, D. A., and Rose, J. M. (2013). Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. *Transportation research part A: policy and practice*, 53:36–52.
6) Carey, M. (1994). A model and strategy for train pathing with choice of lines, platforms, and routes. *Transportation Research Part B: Methodological*, 28(5):333–353.
7) Zhou, X. and Zhong, M. (2007). Single-track train timetabling with guaranteed optimality: Branch-and-bound algorithms with enhanced lower bounds. *Transportation Research Part B: Methodological*, *41*(3), 320-341.
8) Niu, H. and Zhou, X. (2013). Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. *Transportation Research Part C: Emerging Technologies*, 36:212‑230.
9) Barrena, E., Canca, D., Coelho, L. C., and Laporte, G. (2014). Single-line rail rapid transit timetabling under dynamic passenger demand. *Transportation Research Part B: Methodological*, 70:134–150.
10) Niu, H., Zhou, X., and Gao, R. (2015). Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints. *Transportation Research Part B: Methodological*, 76: 117–135.
11) Wang, Y., Tang, T., Ning, B., van den Boom, T. J., and De Schutter, B. (2015). Passengerdemands-oriented train scheduling for an urban rail transit network. *Transportation Research Part C: Emerging Technologies*, 60: 1–23.
12) Robenek, T., Maknoon, Y., Azadeh, S. S., Chen, J., and Bierlaire, M. (2016). Passenger centric train timetabling problem. *Transportation Research Part B: Methodological*, 89:107–126.
13) Cats, O., West, J., & Eliasson, J. (2016). A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. *Transportation Research Part B: Methodological*, 89, 43-57.
14) Shi, J., Yang, L., Yang, J., and Gao, Z. (2018). Service-oriented train timetabling with collaborative passenger flow control on an oversaturated metro line: An integer linear optimization approach. *Transportation Research Part B: Methodological*, 110: 26‑59.
15) Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review*, pages 251‑260.
16) Henderson, J. V. (1974). Road congestion: a reconsideration of pricing theory. *Journal of Urban Economics*, 1(3): 346–365.
17) Hurdle, V. (1981). Equilibrium flows on urban freeways. *Transportation Science*, 15(3): 255–293.
18) Hendrickson, C. and Kocur, G. (1981). Schedule delay and departure time decisions in a deterministic model. Transportation science, 15(1): 62–77.
19) Smith, M. J. (1984). The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation science*, 18(4): 385–394.
20) Daganzo, C. F. (1985). The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation science*, 19(1): 29–37.
21) De Palma, A., Ben-Akiva, M., Lefevre, C., and Litinas, N. (1983). Stochastic equilibrium model of peak period traffic congestion. *Transportation Science*, 17(4): 430–453.
22) Ben-Akiva, M., De Palma, A., and Kanaroglou, P. (1986). Dynamic model of peak period traffic congestion with elastic arrival rates. *Transportation Science*, 20(3): 164–181.
23) Mahmassani, H. S. and Chang, G.-L. (1986). Experiments with departure time choice dynamics of urban commuters. *Transportation Research Part B: Methodological*, 20(4): 297–320.
24) Kraus, M. and Yoshida, Y. (2002). The commuter's time-of-use decision and optimal pricing and service in urban mass transit. *Journal of Urban Economics*, 51(1): 170–195.
25) Tian, Q., Huang, H.-J., and Yang, H. (2007). Equilibrium properties of the morning peakperiod commuting in a many-to-one mass transit system. *Transportation Research Part B: Methodological*, 41(6): 616–631.

26) Gonzales, E. J., & Daganzo, C. F. (2012). Morning commute with competing modes and distributed demand: user equilibrium, system optimum, and pricing. *Transportation Research Part B: Methodological*, *46*(10), 1519-1534.
27) Trozzi, V., Gentile, G., Bell, M. G., & Kaparias, I. (2013). Dynamic user equilibrium in public transport networks with passenger congestion and hyperpaths. *Transportation Research Part B: Methodological*, *57*, 266-285.
28) De Palma, A., Kilani, M., and Proost, S. (2015). Discomfort in mass transit and its implication for scheduling and pricing. *Transportation Research Part B: Methodological*, 71:1‑18.
29) De Palma, A., Lindsey, R., and Monchambert, G. (2017). The economics of crowding in rail transit. *Journal of Urban Economics*, 101:106–122.
30) Seo, T., Wada, K., and Fukuda, D. (2017). A macroscopic and dynamic model of urban rail transit with delay and congestion. *In TRB 96th Annual Meeting Compendium of Papers*.
31) Geroliminis, N. and Levinson, D. M. (2009). Cordon pricing consistent with the physics of overcrowding. In *Transportation and Traffic Theory 2009: Golden Jubilee*, pages 219‑240. Springer.
32) Yang, H. and Tang, Y. (2018). Managing rail transit peak-hour congestion with a fare-reward scheme. *Transportation Research Part B: Methodological*, 110:122‑136.
33) Newell, G. F. (2002). A simplified car-following theory: a lower order model. *Transportation Research Part B: Methodological*, 36(3):195‑205.
34) Edie, L. C. (1963). Discussion of traffic stream measurements and definitions. Port of New York Authority.
35) Arnott, R., De Palma, A., and Lindsey, R. (1990). Economics of a bottleneck. *Journal of urban economics*, 27(1):111‑130.
36) Laih, C.-H. (1994). Queueing at a bottleneck with single- and multi-step tolls. *Transportation Research Part A: Policy and Practice*, 28(3):197‑208.
37) Lindsey, C. R., Van den Berg, V. A., and Verhoef, E. T. (2012). Step tolling with bottleneck queuing congestion. *Journal of Urban Economics*, 72(1):46–59.
38) Yang, H. and Tang, Y. (2018). Managing rail transit peak-hour congestion with a fare-reward scheme. *Transportation Research Part B: Methodological*, 110:122‑136.
39) Zhang, J., Yand, H., Lindsey, R., and Li, X. (2020). Modeling and managing congested transit service with heterogeneous users under monopoly. *Transportation Research Part B: Methodological*, 132:249–266.