Delft University of Technology

Delft Center for Systems and Control

Technical report 14-017

Efficient bi-level approach for urban rail transit operation with stop-skipping*

Y. Wang, B. De Schutter, T.J.J. van den Boom, B. Ning, and T. Tang

If you want to cite this report, please use the following reference instead:
Y. Wang, B. De Schutter, T.J.J. van den Boom, B. Ning, and T. Tang, "Efficient bi-level approach for urban rail transit operation with stop-skipping," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2658–2670, Dec. 2014.

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.24.73 (secretary)
URL: https://www.dcsc.tudelft.nl

Efficient Bi-Level Approach for Urban Rail Transit Operation with Stop-Skipping

Yihui Wang, Bart De Schutter, Ton J.J. van den Boom, Bin Ning, and Tao Tang

***Abstract*—The train scheduling problem for urban rail transit systems is considered with the aim of minimizing the total travel time of passengers and the energy consumption of the trains. We adopt a model-based approach where the model includes the operation of trains at the terminus and at the stations. In order to adapt the train schedule to the origin-destination dependent passenger demand in the urban rail transit system, a stop-skipping strategy is adopted to reduce the passenger travel time and the energy consumption. An efficient bi-level optimization approach is proposed to solve this train scheduling problem, which actually is a mixed integer nonlinear programming problem. The performance of the new efficient bi-level approach is compared with the existing bi-level approach. In addition, we also compare the stop-skipping strategy with the all-stop strategy. The comparison is performed through a case study inspired by real data from the Beijing Yizhuang line. The simulation results show that the efficient bi-level approach and the existing bi-level approach have a similar performance but the computation time of the efficient bi-level approach is around one magnitude smaller than that of the bi-level approach.**

***Index Terms*—train scheduling, urban rail transit, stop-skipping, bi-level optimization, threshold, limiting search space**

I. Introduction

With the increasing passenger demand in large cities (e.g., New York, Tokyo, and Beijing), urban rail transit systems play an increasing significant role for the efficiency and sustainability for the overall transportation process [1]. The operation of trains in several urban rail transit systems is characterized by a high frequency, where the minimum headway between two successive trains is usually 2 to 5 minutes, which could even be further reduced to 90 s with the development of the advanced signaling systems [2]. When trains are operated with such a high frequency, the scheduling of trains based on the passenger demand becomes more and more important for passenger satisfaction and for the reduction of operation costs. The passenger satisfaction can be characterized by the waiting times, in-vehicle times, and the number of transfers, while the operation costs are determined by the number of train services and the energy consumption of train operations.

A. Train scheduling without stop-skipping

Many researchers have explored the train scheduling problem for urban rail transit systems. In 1980, Cury et al. [3] presented a methodology to generate optimal schedules for metro lines based on a model of the train movements and passenger behavior. The performance index includes passenger delay, passenger comfort, and the efficiency of the operation of trains. The resulting nonlinear scheduling problem was recast into many sub-problems by Lagrangian relaxation and then solved in a hierarchical manner [3]. Since the convergence rate of the hierarchical decomposition algorithm can be quite poor in some cases, Assis and Milani [4] proposed a model predictive control algorithm based on linear programming to optimize the train schedule. The algorithm proposed in [4] can effectively generate train schedules for the whole day. In addition, Kwan and Chang [5] applied a heuristic-based evolutionary algorithm to solve the train scheduling problem, where the operation costs and the passenger dissatisfaction are included in the performance index. From the energy-efficiency point of view, Yang et al. [6] proposed a cooperative scheduling approach to optimize the train schedule so that regenerative energy can be used by nearby accelerating trains. Su et al. [7] presented an integrated timetable where the train schedule and speed profiles for trains are optimized at the same time. Furthermore, a demand-oriented timetable design is proposed in [8], where the optimal train frequency and the capacity of trains are first determined and then the schedule of trains are optimized.

B. Train scheduling with stop-skipping

However, since the origin-destination (O-D) passenger demand varies significantly along the urban rail transit line and the time of the day, e.g., some stations (e.g., those in the central business district) have a relatively large number of passengers boarding and alighting and others may have few passengers. Fixed all-stop train schedules cannot efficiently satisfy such an O-D demand pattern. Therefore, we consider the train scheduling problem with stop-skipping for urban rail transit systems in this paper. As demonstrated in [9], [10], the stop-skipping strategy can reduce the passenger travel time and the operation cost of rail transit operators. The skip-stop operation was first developed for the Chicago metro system in 1947 [9]. Now, the SEPTA line in Philadelphia, Helsinki commuter rail, and the metro system in Santiago, Chile apply the stop-skipping train schedule in practice. They apply a static stop-skipping strategy, i.e., the A/B skip-stop strategy, where stations are divided into three types: A, B, and AB; A train services stop at A stations and AB stations, while B train services stop at B stations and AB stations. Major stations are usually labeled with the type AB; so all trains stop there [10].

Y. Wang, B. De Schutter, and T.J.J. van den Boom are with the Delft Center for Systems and Control, Delft University of Technology, The Netherlands, email: {yihui.wang, b.deschutter, a.j.j.vandenboom}@tudelft.nl.

B. Ning and T. Tang are with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China, email: {ttang, bning}@bjtu.edu.cn.

The transit operators provide the stop-skipping information to passengers via panels at platforms and announcements in the trains. The Santiago metro operator stated that passengers adapt to the stop-skipping strategy quickly [9]. In this paper, we propose a dynamic stop-skipping strategy, where the stop-skipping stations for each train are not fixed, but are optimized according to the passenger demand. With the help of screens and announcements at stations, on-board displays and announcements, and personal digital assistant (PDA) devices, passengers can obtain the required information and adapt to the dynamic stop-skipping strategy. In the long run, both the passengers and the rail operator will benefit from stop-skipping. The rail operator can benefit from the shorter cycle times, increased operating speed, and less energy consumption. For most of the passengers, the travel time is shortened and the on-board environment will be better, i.e., lower train occupation. However, the passengers at the skipped stations may experience a longer waiting time and thus possibly a longer total travel time. Therefore, the skipping of trains at stations should be carefully coordinated to benefit passengers. For example, additional constraints can be considered when scheduling the trains, such as: two successive trains should not skip the same station. In this way, the waiting time of passengers can be limited to an acceptable value. Therefore, an efficient solution approach for optimizing the stop-skipping schedule is necessary.

Elberlein [11] formulated the stop-skipping problem as mixed integer nonlinear programming problem, but stop-skipping strings, defined as a collection of consecutive stations, were optimized rather than making the stop-skipping decision for each station. Fu et al. [12] represented the stop-skipping of trains at stations as binary variables and obtained a mixed integer nonlinear programming problem, which was solved using an exhaustive approach. Lee [10] obtained the optimal train schedule using a genetic algorithm. More specifically, first the stop-skipping schedule was predefined for the trains in the urban transit systems. Then the genetic algorithm was applied to find the best combination of these stop-skipping trains and the all-stop trains. In addition, only the travel time of the passengers was considered in [10] and the operation cost of rail transit operation was not considered. In [13], the stop-skipping model allow trains to skip all the stations and a bi-level optimization approach was proposed to solve the train scheduling problem, which is also a mixed integer nonlinear programming problem.

C. Our contributions

The current paper extends the previous research in the following aspects:

- Compared with our previous model in [13], the model here is more compact and includes the operation of trains in the terminus.
- Compared with [10]–[12], both the passenger travel time (including waiting time and in-vehicle time) and the operation costs are taken into account. The operation costs are determined by the number of train services during the scheduling time period and the energy consumption of these trains.

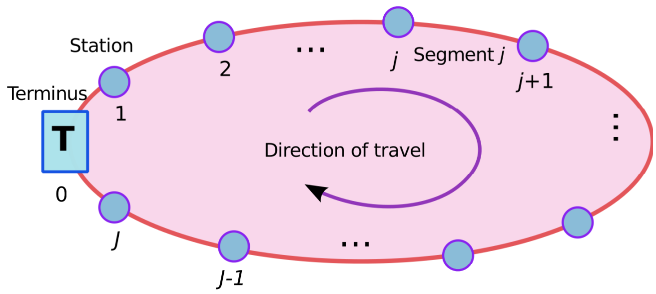


Fig. 1. Illustration of the urban rail transit line

- The stop-skipping model formulated in this paper allows trains to skip the stations specified in a so called skipping set, which could ultimately even involve all the trains and all the stations (except the terminus).
- An efficient bi-level optimization approach is proposed to solve the mixed integer nonlinear programming problem and to obtain the train schedule with stop-skipping for trains.

D. Structure of the paper

This paper is structured as follows. Section II formulates the operation of trains at the terminus, at stations, and in between stations, the passenger demand characteristics, and the passenger/vehicle interaction. Section III describes the multi-objective cost function of the train scheduling problem with stop-skipping. Furthermore, there we also discuss how to solve the scheduling problem in a rolling horizon way and how to define the initial conditions for the scheduling problem. Section IV proposes an efficient bi-level optimization approach for the train scheduling problem with stop-skipping. Section V illustrates the performance of the proposed solution approach with a case study. Finally, Section VI concludes the paper.

II. Model formulation

In this section, the operation of trains and the passenger characteristics for the train scheduling problem with stop-skipping will be formulated.

A. Notation for stations and trains

This paper considers an urban rail transit line as shown in Figure 1, where the terminus and stations in the line are numbered increasingly. Let J denote the total number of stations (terminus not included). The index of the terminus is set equal to 0. The track section between station j and station $j+1$ is denoted as segment j. The scheduling time period for the train scheduling problem is denoted as $[t_0, t_{\text{end}}]$. In order to distinguish the different running cycles of the physical trains, so called train services are introduced, where the train service number in a unique way identifies a train and its current cycle. After the arrival of a physical train at the terminus, its service number will be augmented when the train departs. More specifically, the transit line has I physical trains in total, which are numbered as train $1, 2, \ldots, I$. However, the service number of trains increases with the cycle of the operation of trains. During the scheduling period, the service number of trains is $1, 2, \ldots, I, I+1, I+2, \ldots, 2I, \ldots, N_{\text{cyc}}I$, where N_{cyc} is number of the cycles of the operation of trains for the given time period $[t_0, t_{\text{end}}]$. The service number of a train is increased with I when it departs from the terminus. Therefore, train service i corresponds to physical train $[(i-1) \mod I] + 1$. For the sake of simplicity, we use " train i" as a short-hand for " train service i" from now on.

B. Assumptions

We make the following assumptions for the terminus and the stations:

A.1 Multiple trains can be present at terminus 0, which has a maximum capacity C_0^{ter}. In addition, the trains in terminus 0 will depart from the terminus in a first-in-first-out manner.

A.2 Station j for $j \in \{1, 2, \ldots J\}$ can only accommodate one train at a time and no overtaking can occur at any point of the line.

A.3 Trains can skip some stations in the urban transit line, where we define the skipping set $S = \{(i, j)|$ train i may potentially skip station $j\}$.

A.4 In view of pre-announcement for passengers about the stop-skipping schedule, there is a detailed stop-skipping information displayed in the station and/or the urban rail transit operator provides this information to passengers through mobile devices.

A.5 The operation of trains only consists of three phases: the acceleration phase, the speed holding phase, and the deceleration phase. Moreover, the acceleration and the deceleration are taken to be fixed constants.

A.6 Each passenger only takes one train to arrive at his/her destination, i.e., the transfer between different trains along the line is not allowed.

Assumption A.1 can be motivated as follows: multiple trains can present at the terminus since we assume that there are multiple track sections in the terminus. Furthermore, a first-in-first-out manner for trains in terminus is not difficult to realize in practice, since it depends on the dispatching of trains in the terminus and it is a matter of renumbering of trains. Assumption A.2 generally holds for most urban transit systems, which are usually operated in this way [14], [15]. Even though the stop-skipping strategy is not yet widely used in urban rail transit networks throughout the world, there are several lines which apply it as mentioned before, e.g., the SEPTA line in Philadelphia. Therefore, Assumption A.3 is possible in practice. With the development of technologies, there are already many solutions for the mobile device to provide transportation information for passengers. In addition, the almost all the stations have screens to display travel information to passengers at platforms. Hence, Assumption A.4 is reasonable. In order to simplify the operation model for the trains, the detailed dynamics are not included in the model formulation, but only the three phases mentioned in Assumption A.5 are considered. However, once the running time between two stations are fixed, a more accurate speed profile for the operation of trains can be calculated as a lower level control problem (see [16], [17] for more information). Since in Assumption A.2 we assume that no overtaking can happen at any point of the line, the transfer between different trains for passengers is useless. Therefore, it is reasonable to assume that they will wait at the origin for the right train to get to their destination in Assumption A.6.

C. Operation of a train

The aim of this section is to derive the model equations for the operation of trains in a urban transit line. The operation of trains is controlled through a multi-layer control framework. This paper focus on the train scheduling. In the scheduling layer, we use an online model-based approach; this means the model needs to be simulated repeatedly. Hence, in order to obtain a balanced trade-off between the accuracy and the computation speed, we use a macroscopic model for the scheduling. The detailed train dynamics, the position of block signals, the detection of trains, etc. can then be taken into account by the lower-level control layer. We will first formulate the operation of trains at the terminus and then at the stations.

1) Operation of trains in the terminus: A train can depart from the terminus only after it has arrived. Moreover, the train number is increased with I when it departs from the terminus. So we have

$$d_{i,0} \geq a_{i-I,0} + \tau_{0,\min}, \tag{1}$$

where $d_{i,0}$ is the departure time of train i at terminus 0, a_{i-I} is the arrival time of train $i-I$, and $\tau_{0,\min}$ is the minimum dwell time for the trains at terminus 0. The minimum dwell time could be equal to the minimum turn-around time or the minimum running time at a terminus. In addition, there is no upper bound for the dwell time of trains at terminus. Since there are multiple tracks in terminus to accommodate trains, the running distance for trains between stations and the terminus varies and depends on the route setting in the terminus. However, the layout of the terminus and the scheduling of trains in terminus are out of the scope of this paper. Here, we assume an average distance s_0 for trains running between terminus 0 and station 1 and an average distance s_J for the trains running between station J and terminus 0. The arrival time of train i at terminal 0 is then be written as

$$a_{i,0} = d_{i,J} + r_{i,J}, \tag{2}$$

where $r_{i,J}$ is the running time on segment J.

The headway constraints in terminus 0 can be formulated as

$$d_{i,0} \geq d_{i-1,0} + h_{0,\text{dep}}, \tag{3}$$

where $h_{0,\text{dep}}$ is the minimum departure headway at terminus 0. In addition, the minimum arrival headway at terminus 0 should also be taken into account, which can be formulated as

$$a_{i,0} \geq a_{i-1,0} + h_{0,\text{arr}}, \tag{4}$$

where $h_{0,\text{arr}}$ is the minimum arrival headway at terminus 0.

Remark. *If the departure of trains in the terminal station is affected by the arrival of other trains, then the minimum* *headway constraints between the arrival and departure of trains should also be considered, which can be formulated as:*

$$|d_{i,0} - a_{i',0}| \geq h_{0,\text{dep-arr}},$$

for all $i, i' \in \{1, 2, \ldots, N_{cyc}I\}$ *with* $i \neq i'$.

As mentioned in assumption A.2, the capacity of terminus 0 is C_0^{ter}. Therefore, at any time t the number of trains in terminus 0 should be less than the capacity, which can be formulated as

$$\sum_{i \in S_{\text{trains}}} \mathfrak{I}(a_{i,0} \leq t) - \sum_{i \in S_{\text{trains}}} \mathfrak{I}(t \geq d_{i,0}) \leq C_0^{\text{ter}}, \quad \text{for all } t$$

where S_{trains} is the set of trains considered in the scheduling problem and the indicator function $\mathfrak{I}(\cdot)$ is defined as

$$\mathfrak{I}(x) = \begin{cases} 1 & \text{if } x \text{ is true,} \\ 0 & \text{if } x \text{ is false.} \end{cases}$$

The number of trains in the terminus only increases when a train arrives at the terminus. Therefore, we should only check the capacity constraint when a train arrives at a terminus; so the constraints can be reformulated as

$$\sum_{i \in S_{\text{trains}}} \mathfrak{I}(a_{i,0} \leq a_{\ell,0}) - \sum_{i \in S_{\text{trains}}} \mathfrak{I}(a_{\ell,0} \geq d_{i,0}) \leq C_0^{\text{ter}}, \tag{5}$$

for each $\ell \in S_{\text{trains}}$.

Remark. *This formulation of the operation of trains at a single terminus can be easily extended for the transit lines with multiple termini.*

2) Operation of trains at stations: In assumption A.3, we know that trains can skip some stations. If a train skips a station, then the train passes that station without stopping and the dwell time is then equal to zero. Therefore, passengers are not allowed to get off or on the train. A binary variable is introduced to indicate whether a train will stop at a station or not.

$$y_{i,j} = \begin{cases} 1 & \text{if train } i \text{ will stop at station } j, \\ 0 & \text{if train } i \text{ will skip station } j. \end{cases}$$

The departure time $d_{i,j}$ of train i at station j should satisfy

$$d_{i,j} \geq a_{i,j} + y_{i,j}\tau_{i,j,\min} \tag{6}$$

and

$$d_{i,j} \leq a_{i,j} + y_{i,j}\tau_{i,j,\max}, \tag{7}$$

where $a_{i,j}$ is the arrival time of train i at station j, the minimum dwell time $\tau_{i,j,\min}$ is influenced by the number of passengers boarding and alighting from the train (see (25) in Section II-E for more details), and the maximum dwell time $\tau_{i,j,\max}$ is introduced to ensure passenger satisfaction. The arrival time $a_{i,j+1}$ of train i at station $j+1$ can be calculated by

$$a_{i,j+1} = d_{i,j} + r_{i,j}, \tag{8}$$

where $r_{i,j}$ is the running time of train i on segment j. According to the assumption A.5, the running distances of the acceleration phase, the speed holding phase, and the deceleration can be calculated as

$$s_{i,j}^{\text{acc}} = y_{i,j}\frac{v_{i,j}^2}{2a_{i,j}^{\text{acc}}}, \quad s_{i,j}^{\text{dec}} = y_{i,j+1}\frac{v_{i,j}^2}{2a_{i,j}^{\text{dec}}}, \quad s_{i,j}^{\text{hold}} = s_j - s_{i,j}^{\text{acc}} - s_{i,j}^{\text{dec}},$$

where s_j is the length of segment j, $v_{i,j}$ is the speed of the speed holding phase, $a_{i,j}^{\text{acc}}$ and $a_{i,j}^{\text{dec}}$ are the acceleration and deceleration, respectively. Hence, the running time $r_{i,j}$ of train i for segment j is equal to the sum of the acceleration time, the holding time, and the deceleration time, i.e., $r_{i,j} = t_{i,j}^{\text{acc}} + t_{i,j}^{\text{hold}} + t_{i,j}^{\text{dec}}$, where $t_{i,j}^{\text{acc}} = y_{i,j}v_{i,j}/a_{i,j}^{\text{acc}}$, $t_{i,j}^{\text{dec}} = y_{i,j+1}v_{i,j}/a_{i,j}^{\text{dec}}$, and $t_{i,j}^{\text{hold}} = s_{i,j}^{\text{hold}}/v_{i,j}$. The running time can be recast as

$$r_{i,j} = \frac{s_j}{v_{i,j}} + y_{i,j}\frac{v_{i,j}}{2a_{i,j}^{\text{acc}}} + y_{i,j+1}\frac{v_{i,j}}{2a_{i,j}^{\text{dec}}}. \tag{9}$$

If train i skips stations j, i.e. $y_{i,j} = 0$, then the acceleration time on segment j is equal to zero. If train i skips station $j+1$, i.e., $y_{i,j+1} = 0$, then the deceleration time on segment j is equal to zero. Moreover, if train i skips both station j and $j+1$, then both the acceleration time and deceleration time on segment j are equal to zero. Note that the speed $v_{i,j}$ of the holding phase should satisfy

$$v_{i,j} \in [v_{i,j,\min}, v_{i,j,\max}], \tag{10}$$

where $v_{i,j,\min}$ and $v_{i,j,\max}$ are the minimal and maximal running speed for the speed holding phase of train i at segment j, respectively. The maximum running speed is limited by the train characteristics and the condition of the line. The minimum running speed is introduced to ensure the passenger satisfaction since if trains run too slow, the passengers may complain.

The minimum headway is the minimum time interval between two successive trains so that they can enter and depart from a station safely [18]. Due to assumption A.1, a train cannot enter a station until a minimum headway after the preceding train's departure. With the stop-skipping strategy, the minimum headway between two consecutive trains is in fact affected by whether trains stop at or skip a station. We discern four different cases in our model: (1) two consecutive trains, i.e., train $i-1$ and i, stop at station j, (2) train $i-1$ stops at station j and train i skips station j, (3) train $i-1$ skips station j and train i stops at station j, (4) both train $i-1$ and train i skip station j. Let the minimum headways between two consecutive trains for these four cases be denoted as $h_{j,1}$, $h_{j,2}$, $h_{j,3}$, and $h_{j,4}$, respectively. The headway constraint can then be formulated as

$$\begin{aligned} a_{i,j} - d_{i-1,j} \;\geq\; & y_{i-1,j}y_{i,j}h_{j,1} + (1-y_{i-1,j})y_{i,j}h_{j,2} \\ & + y_{i-1,j}(1-y_{i,j})h_{j,3} + (1-y_{i-1,j})(1-y_{i,j})h_{j,4}. \end{aligned} \tag{11}$$

Remark. *The nonlinear headway constraint* (11) *can also be transformed into linear constraints, since the product of two binary stopping variables* $y_{i-1,j}y_{i,j}$ *can be rewritten as linear constraints. Indeed, the product of two binary variables* $\delta_1\delta_2$ *can be replaced by an auxiliary binary variable* $\delta_3 = \delta_1\delta_2$, *which is equivalent to [19]*

$$\begin{cases} -\delta_1 + \delta_3 \leq 0, \\ -\delta_2 + \delta_3 \leq 0, \\ \delta_1 + \delta_2 - \delta_3 \leq 1. \end{cases} \tag{12}$$

Furthermore, since the end of the scheduling time period is t_{end}, all the departure times should be less than t_{end}, i.e.,

$$d_{i,j} \leq t_{\text{end}}, \text{ for all } i \in \{1,2,\ldots,N_{\text{cyc}}I\} \text{ and } j \in \{1,2,\ldots,J\}. \tag{13}$$

D. Passenger characteristics

The relationship between the variables used for describing the passenger characteristics is illustrated in Figure 2. As we can see from Figure 2 (a), the number of waiting passengers $w_{i,j}^{\text{wait}}$ for train i at station j is equal to the sum of the number of waiting passengers $w_{i,j}^{\text{wait}}$ with destination m for all $m \in \{j+1, j+2, \ldots, m, \ldots, J\}$, i.e.,

$$w_{i,j}^{\text{wait}} = \sum_{m=j+1}^{J} w_{i,j,m}^{\text{wait}}. \tag{14}$$

The number of waiting passengers $w_{i,j}^{\text{wait}}$ with destination m can be calculated by

$$w_{i,j,m}^{\text{wait}} = w_{i-1,j,m} + \lambda_{j,m}(d_{i,j} - d_{i-1,j}), \tag{15}$$

where $w_{i-1,j,m}$ is the number of passengers with destination station m remaining at station j immediately after the departure of train $i-1$, $\lambda_{j,m}(d_{i,j} - d_{i-1,j})$ is the number of newly arrived passengers in between the departures of train i and train $i-1$, and $\lambda_{j,m}$ is the passenger arrival rate at station j for passengers with destination m. Note that the passenger arrival rate at the final station J, is assumed to be zero since we only consider one direction of the line.

In Figure 2, the number of passengers alighting from train i at station j is denoted as $n_{i,j}^{\text{alight}}$, which can be computed using

$$n_{i,j}^{\text{alight}} = \sum_{\ell=1}^{j-1} n_{i,\ell,j}^{\text{board}}, \tag{16}$$

where $n_{i,\ell,j}^{\text{board}}$ is the number of passengers that have station j as their destination and have boarded train i at station ℓ, i.e.,

$$n_{i,\ell,j}^{\text{board}} = w_{i,\ell,j}^{\text{wait}} - w_{i,\ell,j}. \tag{17}$$

No passenger will get off the train if train i skips station j because the passengers at upstream stations were informed that train i would not stop at station j, those passengers with station j as destination would not get on train i. The number of passengers who want to board train i at station j and have station m as their destination is denoted as $w_{i,j,m}^{\text{want-to-board}}$. The number of passengers $w_{i,j,m}^{\text{want-to-board}}$ depends on whether train i stops at station j and whether train i stops at station m for $m \in \{j+1, j+2, \ldots, J\}$, i.e.,

$$w_{i,j,m}^{\text{want-to-board}} = y_{i,j} y_{i,m} w_{i,j,m}^{\text{wait}}. \tag{18}$$

So if train i skips station j, i.e., $y_{i,j} = 0$, then no passengers want to board train i, i.e., $w_{i,j}^{\text{want-to-board}} = 0$. If train i stops at station j, i.e., $y_{i,j} = 1$, then the number of passengers who want to board is decided by whether train i stops at their destination m, i.e., $y_{i,m} w_{i,j,m}^{\text{wait}}$. Note that all the trains stop at terminus 0, so $y_{i,0}$ is equal to 1 for $i \in \{1,2,\ldots,I\}$. As shown in Figure 2 (b), the number of passengers $w_{i,j}^{\text{want-to-board}}$ who want to board train i at station j is

$$w_{i,j}^{\text{want-to-board}} = \sum_{m=j+1}^{J} w_{i,j,m}^{\text{want-to-board}}, \text{ for } m \in \{j+1, j+2, \ldots, J\}. \tag{19}$$

The number of passengers on train i immediately after its departure at station j is defined as $n_{i,j}$, which can be computed as

$$n_{i,j} = n_{i,j-1} - n_{i,j}^{\text{alight}} + n_{i,j}^{\text{board}}, \tag{20}$$

where the number of boarding passengers $n_{i,j}^{\text{board}}$ equals the minimum of the number of passengers that want to board train i and the remaining capacity of the train:

$$n_{i,j}^{\text{board}} = \min(n_{i,j}^{\text{remain}}, w_{i,j}^{\text{want-to-board}}). \tag{21}$$

In addition, the number of passengers $n_{i,j}^{\text{board}}$ boarding train i at station j is also equal to

$$n_{i,j}^{\text{board}} = \sum_{m=j+1}^{J} n_{i,j,m}^{\text{board}}.$$

Moreover, the remaining capacity of train i at station j immediately after the alighting process is

$$n_{i,j}^{\text{remain}} = C_{\max} - n_{i,j-1} + n_{i,j}^{\text{alight}}. \tag{22}$$

The number of passengers $w_{i,j}^{\text{left}}$ left by train i depends on whether train i will stop at station j or not. We have the following two cases:

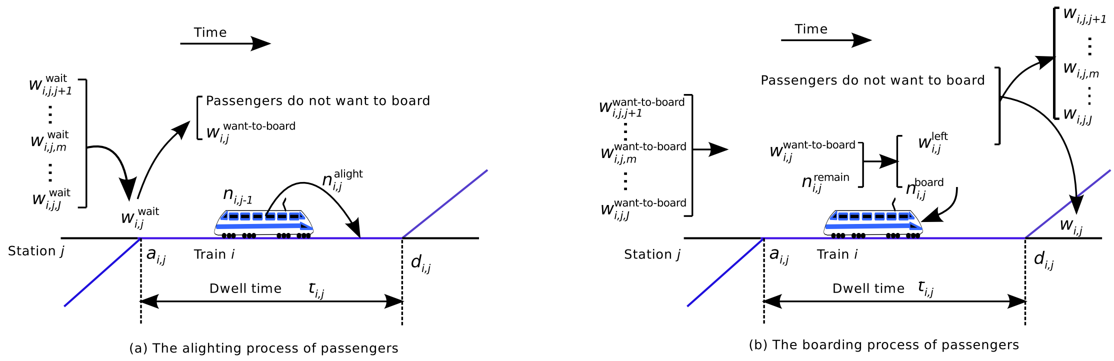
- Train i skips station j, i.e., $y_{i,j} = 0$
  If train i will skip station j, then the number of boarding passengers $n_{i,j}^{\text{board}}$ is equal to zero. All the passengers waiting at station j will then be left by train i.
- Train i will stop at station j, i.e., $y_{i,j} = 1$
  Now consider the case that train i stops at station j. If $w_{i,j}^{\text{want-to-board}} \leq n_{i,j}^{\text{remain}}$, then all the passengers that want to board can get on train i. However, there will be passengers left by train i if $w_{i,j}^{\text{want-to-board}} > n_{i,j}^{\text{remain}}$. The number of passengers who want to board but cannot get on train i at station j immediately after the departure of train i is

  $$w_{i,j}^{\text{left}} = w_{i,j}^{\text{want-to-board}} - \min(n_{i,j}^{\text{remain}}, w_{i,j}^{\text{want-to-board}}) \quad \text{if } y_{i,j} = 1.$$

  In this case, if train i stops station m for $m \in \{j+1, j+2, \ldots, J-1\}$, i.e., $y_{i,m} = 1$, we assume that the number of passengers that have station m as destination and are left by train i is proportional to the number of passengers who want to board. The number of passengers who have destination m and are left by train i can be formulated as

  $$w_{i,j,m} = w_{i,j}^{\text{left}} \frac{w_{i,j,m}^{\text{want-to-board}}}{w_{i,j}^{\text{want-to-board}}} \quad \text{if } y_{i,j} = 1 \text{ and } y_{i,m} = 1.$$

  However, if train i skips station m for $m \in \{j+1, j+2, \ldots, J-1\}$, i.e., $y_{i,m} = 0$, then the number of passengers



(a) The alighting process of passengers

(b) The boarding process of passengers

Fig. 2. The variables for the passenger characteristics

that have station m as destination will not board. So we have

$$w_{i,j,m} = w^{\text{wait}}_{i,j,m}, \quad \text{if } y_{i,j} = 1 \text{ and } y_{i,m} = 0.$$

Hence, the number of passengers who are left by train i and with destination m can be calculated as

$$\begin{aligned} w_{i,j,m} =& y_{i,j}\left(y_{i,m} w^{\text{left}}_{i,j} \frac{w^{\text{want-to-board}}_{i,j,m}}{w^{\text{want-to-board}}_{i,j}} + (1-y_{i,m}) w^{\text{wait}}_{i,j,m} \right) \\ &+ (1-y_{i,j}) w^{\text{wait}}_{i,j,m}. \end{aligned} \tag{23}$$

Furthermore, the total number of waiting passengers at station j immediately after the departure of train i is

$$w_{i,j} = \sum_{m=j+1}^{J} w_{i,j,m}. \tag{24}$$

E. Passenger/vehicle interaction

The minimum dwell time is influenced by the number of passengers boarding and alighting from a train. In addition, the minimum dwell time is also affected by the number of waiting passengers at station: if there are many passengers waiting at the platform, then the boarding process will be slower. In [20], a nonlinear function is given to compute the minimum dwell time:

$$\begin{aligned} \tau_{i,j,\min} =& \min\Bigg(\tilde{\tau}_{\min}, \alpha_{1,\text{d}} + \alpha_{2,\text{d}} n^{\text{alight}}_{i,j} + \alpha_{3,\text{d}} n^{\text{board}}_{i,j} \\ &+ \alpha_{4,\text{d}} \left(\frac{w^{\text{wait}}_{i,j}}{n^{\text{door}}} \right)^3 n^{\text{board}}_{i,j} \Bigg), \end{aligned} \tag{25}$$

where $\tilde{\tau}_{\min}$ is the minimum dwell time predefined by railway operator, $\alpha_{1,\text{d}}$, $\alpha_{2,\text{d}}$, $\alpha_{3,\text{d}}$, and $\alpha_{4,\text{d}}$ are coefficients that can be estimated based on historical data, n^{door} is the number of doors of the train, and $w^{\text{wait}}_{i,j}/n^{\text{door}}$ is the number of passengers waiting at each door. Note that the passengers are assumed to equally distribute over all doors of the train.

III. Real-time train scheduling problem

We first formulate the objective function of the train scheduling problem, which involves the passenger travel time and the energy consumption. Moreover, the waiting time of the passengers who did not travel in the scheduling period is also considered in the objective function. Next, the rolling horizon approach to solve the train scheduling problem is discussed. Furthermore, the initial conditions for the scheduling problem is also defined in this section.

A. Performance criteria

The real-time train scheduling problem is a multi-objective optimization problem, where the objectives could be the energy consumption of the operation of trains, the passenger waiting time, the passenger travel time, and the capacity usage of each train.

The energy consumption of the trains can be calculated by

$$E_{\text{total}} = \sum_{i=1}^{I} \sum_{j=1}^{J} (E^{\text{acc}}_{i,j} + E^{\text{hold}}_{i,j}), \tag{26}$$

where the energy consumption for the acceleration phase $E^{\text{acc}}_{i,j}$ and the holding phase $E^{\text{hold}}_{i,j}$ for train i at segment j can be computed using

$$\begin{aligned} E^{\text{acc}}_{i,j} =& \int_0^{t^{\text{acc}}_{i,j}} \Big((m_\text{e} + n_{i,j} m_\text{p})(a^{\text{acc}}_{i,j} + k_1 + k_2 v(t) + g\sin(\theta_j)) \\ &+ k_3 v^2(t) \Big) v(t) \mathrm{d}t, \end{aligned}$$

$$\begin{aligned} E^{\text{hold}}_{i,j} =& \int_{t^{\text{acc}}_{i,j}}^{t^{\text{acc}}_{i,j} + t^{\text{hold}}_{i,j}} \Big((m_\text{e} + n_{i,j} m_\text{p})(k_1 + k_2 v_{i,j} + g\sin(\theta_j)) \\ &+ k_3 v^2_{i,j} \Big) v_{i,j} \mathrm{d}t. \end{aligned}$$

The total travel time of passengers can be described as a weighted sum of the passenger waiting time and the passenger in-vehicle time

$$t_{\text{total}} = \sum_{i=1}^{I} \sum_{j=1}^{J-1} (\gamma_{\text{wait}} t_{\text{wait},i,j} + t_{\text{in-vehicle},i,j}), \tag{27}$$

where

$$t_{\text{wait},i,j} = w_{i-1,j}(d_{i,j} - d_{i-1,j}) + \frac{1}{2} \sum_{m=j+1}^{J} \lambda_{j,m} (d_{i,j} - d_{i-1,j})^2, \tag{28}$$

and

$$t_{\text{in-vehicle},i,j} = n_{i,j} r_{i,j} + (n_{i,j} - n^{\text{alight}}_{i,j}) \tau_{i,j+1}. \tag{29}$$

Remark. *Since the passengers usually feel that time goes slowly when they are waiting at the platform, a weight larger* *than one can be added to the passenger waiting time in the problem formulation [21].*

In order to spread trains over the entire scheduling time period, we add a penalty term for the waiting time of the passengers left by the last train $N_{\text{cyc}}I$ during the scheduling period:

$$f_{\text{penalty},1} = \mu_{\text{penalty},1} \sum_{j=1}^{J} \left(w_{N_{\text{cyc}}I,j}(t_{\text{end}} - d_{N_{\text{cyc}}I,j}) + \frac{1}{2}\zeta \sum_{m=j+1}^{J} \lambda_{j,m}(t_{\text{end}} - d_{N_{\text{cyc}}I,j})^2 \right), \tag{30}$$

If $\zeta = 1$, the waiting time for the newly arrived passengers between the departure time $d_{N_{\text{cyc}}I,j}$ and the end time t_{end} is also considered. However, if $\zeta = 0$, the waiting time of the newly arrived passengers after the last train is not considered (e.g., the trains coming later will pick up these passengers). But in the latter case, we need to add a penalty term for the arrival time of the last train $N_{\text{cyc}}I$ at the terminus to avoid all the trains operating close to each other at the start of the period $[t_0, t_{\text{end}}]$:

$$f_{\text{penalty},2} = \mu_{\text{penalty},2} |a_{N_{\text{cyc}}I,0} - t_{\text{end}}|. \tag{31}$$

The objective function of the train scheduling problem can be written as

$$f_{\text{opt}} = \gamma_1 \frac{E_{\text{total}}}{E_{\text{total,nom}}} + \gamma_2 \frac{t_{\text{total}}}{t_{\text{total,nom}}} + f_{\text{penalty},1} + f_{\text{penalty},2}, \tag{32}$$

where γ_1 and γ_2 are non-negative weights, and the normalization factors $E_{\text{total,nom}}$ and $t_{\text{total,nom}}$ are "nominal" values of the total energy consumption and the total travel time of passengers, respectively. These nominal values can e.g. be determined by running trains using a feasible initial schedule.

B. Rolling horizon approach and initial conditions

Since the passenger demand varies with the time in a daily operation, the train scheduling problem can be solved in a rolling horizon way, by solving the scheduling problem, e.g., every half an hour, so as to adapt the train schedule to passenger demand in real time. This works as follows. First, the train scheduling problem is solved for some period $[t_0, t_{\text{end}}]$ and the trains will be run according to the resulting optimal schedule. After some period of time h, e.g., half an hour, we will run the optimization process again, but now for the period $[t_0 + h, t_{\text{end}} + h]$ using the known, measured, or estimated states of the system at time $t_0 + h$. Once the new optimal schedule is computed, it is executed for h time units, and next the whole process is repeated again for the period $[t_0 + 2h, t_{\text{end}} + 2h]$ and so on, until the end of the daily operation of the urban rail transit system.

When solving the train scheduling problem in a rolling horizon way, some of the variables will no longer free variables but will have fixed values. Let t_0 be the start time instant of the scheduling period, we now discuss the fixed variables:

- If train i is in terminus 0 at time t_0, i.e., the arrival time $a_{i-I,0}$ of train $i-I$ at terminus 0 will be a known time value before t_0. So $a_{i-I,0}$ is no longer an unknown variable.
- If train i is running on a segment at t_0, we use j_{i,t_0} to denote the segment at which train i is running on at t_0. The departure time $d_{i,j_{i,t_0}}$ of train i at station j_{i,t_0} is a known time value before t_0. In addition, all the departure times, arrival times, and running times before segment j_{i,t_0} are known. Furthermore, the running time $r_{i,j_{i,t_0}}$ on segment j_{i,t_0} is also fixed since we assume that the schedule of a train can only be changed at stations. Therefore, the arrival time of train i at station $j_{i,t_0}+1$ is also known.
- If train i is at station at time t_0, we use j_{i,t_0} to denote the station at which train i is stopping at. The arrival time $a_{i,j_{i,t_0}}$ of train i at station j_{i,t_0} is known. In addition, the departure times, the arrival times, and the running times before station j_{i,t_0} are also known.

Moreover, the stopping variables are fixed for trains that are already on their way to make sure all the passengers on the train can arrive their destinations. The number of passengers on the train and the number of passengers waiting at the platform are also known at time t_0.

The train scheduling model also requires the real-time assessment of the passenger arrival rates in the O-D matrix during the scheduling period. In the case of full state information, the passenger arrival rates can be obtained. However, this is not the case in practice, where we can e.g. use the information collected by the advanced fare collection systems and estimate the passenger arrival rates based on the historical data and the current passenger flows [22].

IV. Efficient solution approach

The resulting real-time scheduling problem is a mixed integer nonlinear programming (MINLP) problem with objective function (32) and constraints (1)-(24). In [13], we have proposed a bi-level optimization approach to solve this optimization problem. However, the computation time of this bi-level optimization method is too long in practice. Therefore, we now propose an efficient bi-level solution approach for the MINLP problem, where the search space of the problem is limited and a threshold method is presented to obtain a good initial solution for the MINLP problem.

A. Bi-level optimization

The free variables in the real-time scheduling problem are the departure time $d_{i,j}$, the holding speed $v_{i,j}$, and the binary variables $y_{i,j}$ for $(i, j) \in S$ (i.e., the set of train i and station j where skipping is possible). The other variables like the number of passengers waiting at stations $w_{i,j}$ and the number of passengers on-board the trains $n_{i,j}$ can be eliminated using the model equation (13)-(23). The bi-level optimization method proposed in [13] consists of two levels of optimization:

- The high-level optimization optimizes the binary variables $y_{i,j}$ for $(i, j) \in S$, where a brute force approach can be used to explore all the combinations for the binary variables in case the size of the problem is small. Alternatively, integer programming approaches, such as genetic algorithms or branch-and-bound methods, can be applied in the high-level optimization.

- For each combination of binary variables, the low-level optimization solves a nonlinear non-convex problem using e.g., multi-start sequential quadratic programming (SQP) algorithm or a pattern search method.

B. Threshold method for obtaining good initial solutions

In order to obtain a good initial solution for the train scheduling problem, we first introduce a threshold function to determine the value of the stopping variable as follows:

$$y_{i,j} = \Im\Big(\big(w_{i,j}^{\text{want-to-board}} \geq \theta_{i,j}^{\text{th,in}}\big) \wedge \big(n_{i,j}^{\text{alight}} \geq \theta_{i,j}^{\text{th,out}}\big)\Big), \quad (33)$$

where $\theta_{i,j}^{\text{th,in}}$ and $\theta_{i,j}^{\text{th,out}}$ are the thresholds, which are free variables and determined by the optimization procedure. In this case, the value of the stopping variables depends on the passenger flows in the urban rail transit line. By introducing the threshold function, we can reformulate the MINLP problem as a real-valued nonlinear programming problem, which can be solved by sequential quadratic programming method. An initial solution of the stopping variables, the departure times, and the holding speeds can be obtained by solving this nonlinear programming problem.

C. Limiting the search space

After the initial solution is obtained using the threshold method, we can limit the search space of the train scheduling problem within a neighborhood of the initial solution to reduce the computation time. For the high-level optimization, the search space of the integer variables $y = \left[y_{1,1}y_{1,2},\ldots,y_{2,1},\ldots,y_{N_{\text{cyc}I},J}\right]^T$ can be limited by the following 1-norm constraint:

$$\|y - y_{\text{init}}\|_1 \leq \chi_0, \quad (34)$$

which means that only a limited number, i.e., χ_0, of binary variables can change their values in the bi-level optimization approach. If χ_0 is small, the search space can be reduced dramatically. The brute force method can be applied for the high-level optimization if χ_0 is chosen as 1 or 2. Otherwise, the genetic algorithm can be applied.

In addition, we can also limit the search space of the departure times and holding speeds as follows:

$$\|d - d_{\text{init}}\| \leq \chi_2 + \chi_3\beta, \quad (35)$$

and

$$\|v - v_{\text{init}}\| \leq \chi_3 + \chi_4\beta, \quad (36)$$

where β is flexibility variable introduced to make sure the optimization problem is always feasible. The objective function (32) is revised as

$$f'_{\text{opt}} = f_{\text{opt}} + \gamma_4\beta, \quad (37)$$

where $\gamma_4\beta$ introduces flexibility in the degree of goal attainment.

TABLE I
PARAMETERS OF THE TRAINS AND THE PASSENGERS

Property	Symbol	Value
Train mass [kg]	$m_{\text{e,i}}$	$199 \cdot 10^3$
Mass of one passenger [kg]	m_{p}	60
Capacity of trains [passenger]	$C_{i,\max}$	1468
Minimum dwell time [s]	$\tilde{\tau}_{\min}$	30
Maximum dwell time [s]	$\tau_{\max}$	150
Coefficients of the minimal dwell time	$\alpha_{1,d}$	4.002
	$\alpha_{2,d}$	0.047
	$\alpha_{3,d}$	0.051
Coefficients of resistance	k_{1i}	0.012
	k_{2i}	$5.049 \cdot 10^{-4}$
	k_{3i}	8.521

V. Case Study

A. Set-up

In order to demonstrate the effectiveness of the proposed model formulation and the performance of the proposed efficient bi-level optimization approach, we consider a cyclic line with 1 terminus and 12 stations following the structure shown in Figure 1. There are 6 physical trains in the urban rail transit line and the schedule for 10 train services will be optimized. This means that the trains for the train services 1-4 will reach the terminal station and then start a new train service during the scheduling period. The new train services are numbered from 7 to 10. The train characteristics and the line data are inspired by the data of Beijing Yizhuang subway line, and are given in Table I and Table II, respectively.

In Table II, station 0 represents terminus 0. The minimum running time in Table II is calculated by taking a fixed acceleration of 0.8 m/s^2 and a fixed deceleration of -0.8 m/s^2; furthermore, the trains are assumed to run at the maximum speed of 80 km/h during the holding phase. The maximum running time is assumed as $r_{i,j,\max} = \zeta r_{i,j,\min}$, where ζ is larger than 1. We have chosen ζ as 1.2 to ensure that the passengers do not complain that the train is too slow. The mass of the train, the mass of one passenger, and the coefficients for the minimum dwell time in (25) are given in Table I. The lower bound for the dwell time predefined by the railway operator is chosen as 30 s. The passenger arrival rates at stations are shown in Table III.

The initial states at time t_0 (chosen as 1300 s for this case study) of the trains are as follows: train 1 and 2 are running to station 8 and 5, respectively. Since we assume that the schedule of a train can only be changed at stations, so the arrival times of these two trains at station 3 and station 2 are fixed and they are 1400 s and 1340 s, respectively. Train 3 are stopping at station 3 and its arrival time is 1270 s. The number of passengers on train 1, 2, and 3 at time t_0 their destination are given in Table IV.

In addition, there are 3 trains stopping at the terminus, so the previous train services finished before t_0. The communication-based train control system (a moving block signaling system) is implemented in Beijing Yizhuang subway line, where the minimum headway between two successive trains is 90 s. In addition, a maximum departure-departure headway is included to ensure the passenger satisfaction, which is chosen as 400 s.

TABLE II
INFORMATION OF THE CYCLIC LINE

Station number	0	1	2	3	4	5	6	7	8	9	10	11	12
Distance to next station [m]	1050	1832	1786	2086	2265	1030	1354	1280	1544	992	1975	2369	1349
Minimal running time [s]	75.0	110.2	108.2	121.7	129.7	74.1	88.7	85.4	97.3	72.4	116.7	134.4	88.5

TABLE III
THE ORIGIN-DESTINATION DEPENDENT PASSENGER ARRIVAL RATES AT STATIONS [PASSENGERS/S]

Station	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0.06	0.30	0.35	0.03	0.18	0.36	0.06	0.34	0.27	0.03	0.12
2	0	0	0.05	0.06	0.02	0.01	0.04	0.02	0.02	0.01	0.02	0.03
3	0	0	0	0.2	0.03	0.27	0.18	0.02	0.25	0.17	0.03	0.36
4	0	0	0	0	0.02	0.25	0.25	0.04	0.22	0.32	0.02	0.34
5	0	0	0	0	0	0.05	0.02	0.01	0.01	0.04	0.01	0.03
6	0	0	0	0	0	0	0.21	0.08	0.24	0.27	0.05	0.39
7	0	0	0	0	0	0	0	0.02	0.35	0.23	0.03	0.34
8	0	0	0	0	0	0	0	0	0.03	0.04	0.02	0.05
9	0	0	0	0	0	0	0	0	0	0.27	0.03	0.39
10	0	0	0	0	0	0	0	0	0	0	0.03	0.35
11	0	0	0	0	0	0	0	0	0	0	0	0.06
12	0	0	0	0	0	0	0	0	0	0	0	0

TABLE IV
THE NUMBER OF PASSENGERS ON TRAIN 1, 2, AND 3 AT TIME t_0 AND THEIR DESTINATIONS

Destination station	1	2	3	4	5	6	7	8	9	10	11	12	Total passengers
Train 1	0	0	0	0	0	0	0	0	131	395	263	132	921
Train 2	0	0	0	0	0	89	33	111	44	333	166	22	798
Train 3	0	0	0	106	100	20	144	216	31	103	144	21	885

Furthermore, the number of passengers waiting at the various stations at t_0 and the destination of these passengers are shown in Table V. The nominal values for the total travel time, the energy consumption, and the waiting time for the passengers who did not travel in the scheduling period is calculated based on a schedule with constant headway, which are $2.278 \cdot 10^7$ s, $7.013 \cdot 10^9$ J, and $1.387 \cdot 10^7$ s, respectively.

B. Simulation and results

The train scheduling problem is solved using the following three approaches:

- All-stop approach: Trains in the scheduling period stop at every station, i.e., there is no stop-skipping at all. In this case, the train scheduling problem is a nonlinear programming problem, which is solved here using the sequential quadratic programming (SQP) method implemented by the fmincon function of Matlab optimization toolbox.
- Bi-level approach with stop-skipping: The train scheduling problem with stop-skipping is a mixed integer nonlinear programming problem, which is solved using the bi-level approach in [13]. A genetic algorithm is applied for the integer optimization of the high level, where the ga function of the global optimization toolbox of Matlab is employed. The nonlinear optimization problem in the lower level is solved using the SQP algorithm of the fmincon function of the Matlab optimization toolbox.
- Efficient bi-level approach with stop-skipping: First, a good initial solution for the mixed integer nonlinear problem is obtained by solving the train scheduling problem with threshold function (33), which is a nonlinear programming problem and which can be solved e.g. using multistart SQP. Next, the resulting solution is used as starting point for solving the full optimization problem, but with the search space of the binary variables limited by the 1-norm constraints (34). The size of the search space varies with the value assigned to χ_0. When χ_0 is equal to 1, we apply a brute force approach for the high-level optimization since the search space of the binary variables is small. When χ_0 is equal to 2 and 3, a genetic algorithm is used to optimize the binary variables. Furthermore, the SQP algorithm is employed by the low-level optimization, where the search space can also be limited.

The train schedules obtained by the all-stop approach, the bi-level approach, and the efficient bi-level approach are shown in Figures 3-8. These train schedules look similar to each other, however, there are some differences between them. In particular, for the all-stop approach (Figure 3) all trains stop at all stations, while several trains skip some stations in the train schedules obtained by the bi-level approach and the efficient bi-level approach (see Figures 4-8). In the train schedule obtained by the bi-level approach shown in Figure 4, trains 4, 6, 7, 8, 9, and 10 skip some stations. More specifically, train 4 skips stations 2, 5, 8, and 11, so the stopping variables of train 4 for these stations are equal to 0 as shown in Table VI[1]. In addition, we can observe that the travel time for trains that skip some stations is smaller than that of the all-stop approach, e.g., train 4 arrives earlier at the terminal station in

[1]Because the stopping variables for other stations are all equal to 1, we only list the stopping variables for stations 2, 5, 8, and 11.

TABLE V
THE NUMBER OF PASSENGERS WAITING AT STATIONS AT t_0 AND THEIR DESTINATIONS

Destination	1	2	3	4	5	6	7	8	9	10	11	12	Total passengers
Station 1	0	8	75	54	65	15	26	32	15	68	33	21	421
Station 2	0	0	19	19	19	14	16	13	15	13	11	14	153
Station 3	0	0	0	29	91	15	45	32	12	41	33	14	312
Station 4	0	0	0	0	57	22	52	43	22	11	24	35	266
Station 5	0	0	0	0	0	13	22	29	26	14	26	24	154
Station 6	0	0	0	0	0	0	25	23	26	5	30	14	123
Station 7	0	0	0	0	0	0	0	23	25	13	27	29	117
Station 8	0	0	0	0	0	0	0	0	25	9	19	18	71
Station 9	0	0	0	0	0	0	0	0	0	24	20	23	67
Station 10	0	0	0	0	0	0	0	0	0	0	21	25	46
Station 11	0	0	0	0	0	0	0	0	0	0	0	27	27
Station 12	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4 (stop-skipping approach) than in Figure 3 (all-stop approach). Trains 1, 2, and 3 have already departed from the terminus at time t_0 and they are thus suppose to stop at all stations. So the stopping variables for these three trains are equal to 1. Figure 5 illustrates an initial train schedule that is obtained by the threshold method. The values of the stopping variables are shown in Table VI, which are different from those obtained via the bi-level approach. For example, train 4 only skips stations 5 and 8 but does not skip station 2 and 11. Based on the initial train schedule shown in Figure 5, we assign the value of χ_0 in (34) as 1, 2, and 3 to vary the search space. Figure 6 shows the train schedule obtained by the efficient bi-level approach with $\chi_0 = 1$, which means that one binary variable can change its value. From Figure 6 and Table VI, we can observe that train 7 skips station 5 when $\chi_0 = 1$, which is different from the initial schedule. Similarly, when $\chi_0 = 2$, the values of two binary variables are changed comparing with the initial solution. As we can observe from Figure 7 and Table VI, train 4 skips station 11 and train 7 skips station 5, while in the initial train schedule shown in Figure 5, train 4 stops at station 11 and train 7 stops at station 5. Figure 8 presents the train schedule obtained by the efficient bi-level approach with $\chi_0 = 3$, where train 4 skips station 11 and both train 7 and train 9 skips station 5 compared with the initial schedule.

A comparison of the performance of these three approaches is illustrated in Table VII, where the values of the objective function, the computation time, the total passenger travel time, the energy consumption of the train schedules, etc. are listed. The relative improvements of the bi-level approach and the efficient bi-level approach with respect to the all-stop approach are given as in Table VIII, which is calculated as

$$x_{\text{relative-difference}} = 1 - \frac{x_{\text{stop-skipping}}}{x_{\text{all-stop}}},$$

where x is the value of the objective function, the computation time, etc. in the table and the stop-skipping involves the solution obtained by the bi-level approach and the efficient bi-level approach. Note that in Table VII, the solution approach has a better performance when the number of passengers finished their trips is larger since more passengers have finished their trip during the scheduling period. For the other terms in Table VII, the performance of the approach is better if these terms have a smaller value. In a similar way, the relative improvement of the number of passengers that finished their trip is smaller means the improvement is better. For the other terms in Table VIII, a bigger value represents better performance.

C. Discussion

For the given case study, the overall performance improvement of the stop-skipping strategy is about 8-12% better compared with the all-stop approach. With the stop-skipping strategy, the total travel time is reduced with 12-15% and the total energy consumption is reduced with 10-15%. The number of passengers that did not travel increases with 7-11%; however, note that the trains coming later will pick up these passengers anyway; so they will not be left at the platform. Since we solving the train scheduling problem in a rolling horizon way, the passengers that did not travel at the current time period will be taken into account in the next period (cf. Section III-B). The efficient bi-level approach yields an acceptable performance when compared with the bi-level approach. However, the computation time of the bi-level approach is about 10 and 4 times longer than that of the efficient bi-level approach with $\chi_0 = 1$ and $\chi_0 = 3$, respectively. More specifically, the computation time of the efficient bi-level approach (with $\chi_0 = 1$) is about half an hour using Matlab on a 64-bit Linux operation system running on a 1.8 GHz Intel Core2 Duo CPU. Hence, if we would use dedicated optimization software written in object code, in combination with a faster processors and parallel processing, the efficient bi-level approach is tractable for real-time application.

VI. CONCLUSIONS

We have considered the train scheduling problem with stop-skipping, where the operation of trains at both the stations and the terminus are included in the model. The passenger characteristics are described based on the origin-destination demands for passengers. Since the resulting train scheduling problem is a mixed integer nonlinear programming problem, an efficient bi-level approach has been proposed, where a threshold method is applied to obtain a good initial solution for

[2] These remaining passengers, i.e., the passengers that did not travel, will be picked up by the trains that arrive later on. The waiting time of those passengers is also included in the objective function. So here the number and the waiting time of these passengers are also given for comparing the performance of these three approaches.

TABLE VI
THE STOPPING VARIABLES OF THE ALL-STOP APPROACH, THE BI-LEVEL APPROACH, AND THE EFFICIENT BI-LEVEL APPROACH AT STATION 2, 5, 8, AND 11 (IF THE STOPPING VARIABLE EQUALS 1, THEN THE TRAIN STOPS AT THE STATION; OTHERWISE, THE TRAIN WILL SKIP THE STATION)

Solution approaches	Solution options	Train	1	2	3	4	5	6	7	8	9	10
Bi-level	–	Station 2	1	1	1	0	1	1	0	0	1	1
		Station 5	1	1	1	0	1	0	0	0	0	0
		Station 8	1	1	1	0	1	0	0	0	1	0
		Station 11	1	1	1	0	1	0	0	0	0	1
Efficient bi-level	Solution (threshold)	Station 2	1	1	1	1	0	0	1	0	0	1
		Station 5	1	1	1	0	0	1	1	0	1	0
		Station 8	1	1	1	0	1	0	1	1	1	1
		Station 11	1	1	1	1	1	0	1	0	1	0
	Solution ($\chi_0 = 1$)	Station 2	1	1	1	1	0	0	1	0	0	1
		Station 5	1	1	1	0	0	1	0	0	1	0
		Station 8	1	1	1	0	1	0	1	1	1	1
		Station 11	1	1	1	1	1	0	1	0	1	0
	Solution ($\chi_0 = 2$)	Station 2	1	1	1	1	0	0	1	0	0	1
		Station 5	1	1	1	0	0	1	0	0	1	0
		Station 8	1	1	1	0	1	0	1	1	1	1
		Station 11	1	1	1	0	1	0	1	0	1	0
	Solution ($\chi_0 = 3$)	Station 2	1	1	1	1	0	0	1	0	0	1
		Station 5	1	1	1	0	0	1	0	0	0	0
		Station 8	1	1	1	0	1	0	1	1	1	1
		Station 11	1	1	1	0	1	0	1	0	1	0

TABLE VII
PERFORMANCE COMPARISON OF THE ALL-STOP APPROACH, THE BI-LEVEL APPROACH, AND THE EFFICIENT BI-LEVEL APPROACH (χ_0 IS THE NUMBER OF BINARY VARIABLES, THE VALUE OF WHICH CAN BE MODIFIED)

Solution approach	All-stop	Bi-level	Efficient bi-level			
			Initial solution (threshold)	Solution ($\chi_0 = 1$)	Solution ($\chi_0 = 2$)	Solution ($\chi_0 = 3$)
Objective value	2.917	2.617	2.708	2.681	2.661	2.653
Computation time [s]	$4.464 \cdot 10^2$	$1.672 \cdot 10^4$	$1.194 \cdot 10^3$	$1.705 \cdot 10^3$	$2.352 \cdot 10^3$	$4.567 \cdot 10^3$
Energy consumption [J]	$5.888 \cdot 10^9$	$4.994 \cdot 10^9$	$5.369 \cdot 10^9$	$5.301 \cdot 10^9$	$5.220 \cdot 10^9$	$5.111 \cdot 10^9$
Number of passengers finished their trips [passenger]	$2.603 \cdot 10^4$	$2.451 \cdot 10^4$	$2.500 \cdot 10^4$	$2.505 \cdot 10^4$	$2.482 \cdot 10^4$	$2.473 \cdot 10^4$
Number of passengers that did not travel [passenger]	$1.324 \cdot 10^4$	$1.476 \cdot 10^4$	$1.427 \cdot 10^4$	$1.422 \cdot 10^4$	$1.445 \cdot 10^4$	$1.455 \cdot 10^4$
Travel time for passengers finished their trips [s]	$1.944 \cdot 10^7$	$1.648 \cdot 10^7$	$1.724 \cdot 10^7$	$1.706 \cdot 10^7$	$1.680 \cdot 10^7$	$1.683 \cdot 10^7$
Waiting time of passengers that did not travel [2] [s]	$1.031 \cdot 10^7$	$1.270 \cdot 10^7$	$1.192 \cdot 10^7$	$1.188 \cdot 10^6$	$1.227 \cdot 10^6$	$1.240 \cdot 10^6$

TABLE VIII
RELATIVE IMPROVEMENT WITH RESPECT TO THE ALL-STOP APPROACH OF THE BI-LEVEL APPROACH, AND THE EFFICIENT BI-LEVEL APPROACH (χ_0 IS THE NUMBER OF BINARY VARIABLES, THE VALUE OF WHICH CAN BE MODIFIED)

Solution approach	Bi-level	Efficient bi-level			
		Initial solution (threshold)	Solution ($\chi_0 = 1$)	Solution ($\chi_0 = 2$)	Solution ($\chi_0 = 3$)
Objective value	12.01%	7.16%	8.08%	8.77%	9.05%
Computation time [s]	−36.46	−1.68	−2.82	−4.27	−9.23
Energy consumption [J]	15.18%	8.81%	9.97%	11.35%	13.20%
Number of passengers finished their trips [passenger]	5.83%	3.94%	3.77%	4.64%	5.01%
Number of passengers that did not travel [passenger]	−11.46%	−7.75%	−7.41%	−9.12%	−9.85%
Travel time for passengers finished their trips [s]	15.19%	11.32%	12.24%	13.55%	13.41%
Waiting time of passengers that did not travel [s]	−23.26%	−15.64%	−15.29%	−19.04%	−20.35%

the full problem and where the search space for the variables can be limited to enhance the efficiency. For a case study, the efficient bi-level approach with a limited search space provided the best solution within the time that is typically available for the computations (e.g., half an hour). In particular, the overall performance improved with about 8-12% compared to the all-stop approach.

The approaches proposed in this paper can be easily extended to other urban rail transit lines. The parameters for train dynamics are usually available for the rail operators. The origin-destination-dependent (OD-dependent) passenger arrival rates can be estimated from the data obtained by the automatic fare collection system. The initial states of the system can be obtained by detection equipments at stations and on-board trains. In our future work, we will perform additional case studies for larger real-life networks where we will also include real-life measurements (e.g., demand profiles). Furthermore, we will investigate other solution approaches to solve the mixed integer nonlinear programming problem efficiently, especially for cases with a large number of trains and stations. In addition, we will investigate the effect of more detailed models (including short turns, the distribution

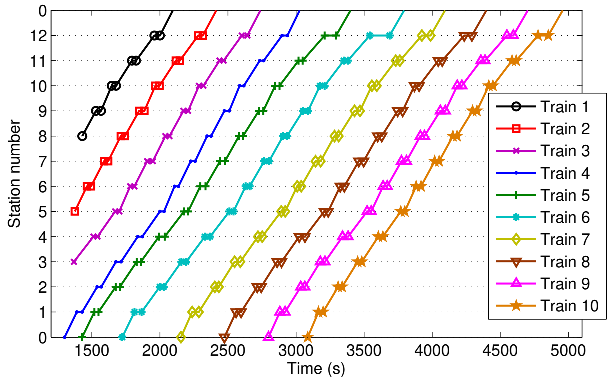


Fig. 3. The train schedule obtained by the all-stop approach

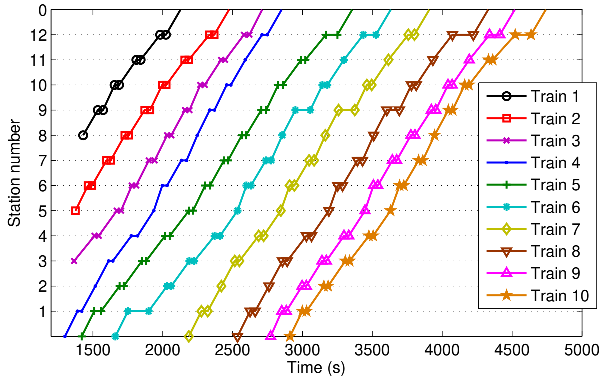


Fig. 4. The train schedule obtained by the bi-level approach

of on-board passengers and waiting passengers at platforms, etc.) on the trade-off between performance and computational complexity. Moreover, robust train scheduling is important in practice since there are stochastic disturbances during the operation of trains [23], [24]. In future work, we will consider the travel time uncertainty, dwell time uncertainty, etc. in the train scheduling and investigate robust train scheduling.

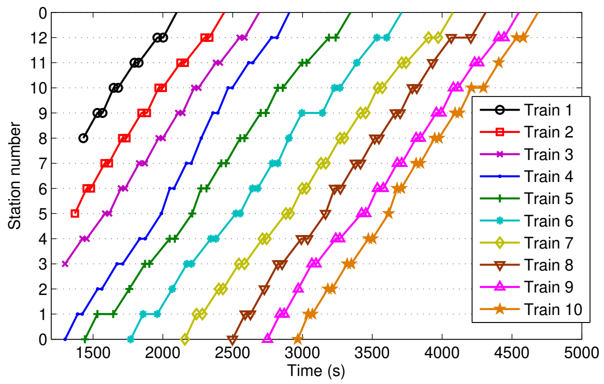


Fig. 5. The train schedule obtained by the threshold method

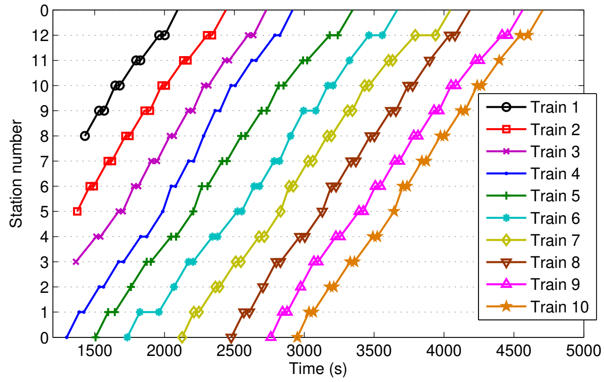


Fig. 6. The train schedule obtained by the efficient bi-level approach with $\chi_0 = 1$

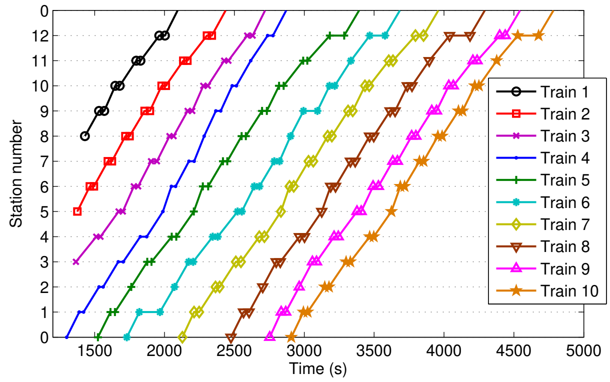


Fig. 7. The train schedule obtained by the efficient bi-level approach with $\chi_0 = 2$

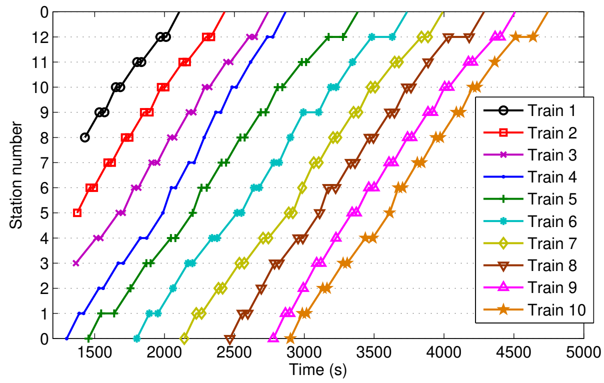


Fig. 8. The train schedule obtained by the efficient bi-level approach with $\chi_0 = 3$

ACKNOWLEDGMENT

Research supported by the Dutch Technology Foundation STW project “Model-predictive railway traffic management” (11025), the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement no. 257462 HYCON2 Network of Excellence, the Beijing Municipal Science and Technology Commission (Contract No. D111100000411001), and the State Key Laboratory of Rail Traffic Control and Safety (Contract No. RCS2010ZZ003).

REFERENCES

[1] H. Dong, B. Ning, Y. Chen, X. Sun, D. Wen, Y. Hu, and R. Ouyang, “Emergency management of urban rail transportation based on parallel systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 627–636, Jun. 2013.

[2] H. Takeuchi, C. Goodman, and S. Sone, “Moving block signaling dynamics: performance measures and re-starting queued electric trains,” *IEE Proceedings - Electric Power Applications*, vol. 150, no. 4, pp. 483–492, Jul. 2003.

[3] J. E. Cury, F. A. C. Gomide, and M. J. Mendes, “A methodology for generation of optimal schedules for an underground railway systems,” *IEEE Transaction on Automatic Control*, vol. 25, no. 2, pp. 217–222, Apr. 1980.

[4] W. O. Assis and B. E. A. Milani, “Generation of optimal schedules for metro lines using model predictive control,” *Automatica*, vol. 40, no. 8, pp. 1397–1404, Aug. 2004.

[5] C. Kwan and C. S. Chang, “Application of evolutionary algorithm on a transportation scheduling problem - the mass rapid transit,” in *Proceedings of the 2005 IEEE Congress on Evolutionary Computation*, vol. 2, Edinburgh, Scotland, Sep. 2005, pp. 987–994.

[6] X. Yang, X. Li, Z. Gao, H. Wang, and T. Tang, “A cooperative scheduling model for timetable optimization in subway systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 438–447, Mar. 2013.

[7] S. Su, X. Li, T. Tang, and Z. Gao, “A subway train timetable optimization approach based on energy-efficient operation strategy,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 883–893, Jun. 2013.

[8] T. Albrecht, “Automated timetable design for demand-oriented service on suburban railways,” *Public Transport*, vol. 1, no. 1, pp. 5–20, May 2009.

[9] M. Freyss, R. Giesen, and J. C. M. noz, “Continuous approximation for skip-stop operation in rail transit,” *Transportation Research Part C*, vol. 36, pp. 419–433, Nov. 2013.

[10] Y. J. Lee, “Mathematical modeling for optimizing skip-stop rail transit operation strategy using genetic algorithm,” Morgan State University, Department of Transportation and Urban Infrastructure Studies, Baltimore, US, Tech. Rep., 2012.

[11] X. J. Elberlein, *Real-time Control Strategies in Transit Operations: Models and Analysis*. Massachusetts Institute of Technology, Cambridge, USA: Ph.D. thesis, 1995.

[12] L. Fu, Q. Liu, and P. Calamai, “Real-time optimization model for dynamic scheduling of transit operations,” *Transportation Research Record*, vol. 1857, pp. 48–55, Jan. 2003.

[13] Y. Wang, B. De Schutter, T. van den Boom, B. Ning, and T. Tang, “Origin-destination dependent train scheduling problem with stop-skipping for urban rail transit systems,” in *Proceedings of the 93nd Annual Meeting of the Transportation Research Board*, Washington, DC, Jan. 2014.

[14] X. J. Elberlein, N. H. M. Wilson, and D. Bernstein, “The holding problem with real-time information available,” *Transportation Science*, vol. 35, no. 1, pp. 1–18, Feb. 2001.

[15] H. Wang, F. Schmid, L. Chen, C. Roberts, and T. Xu, “A topology-based model for railway train control systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 819–827, Jun. 2013.

[16] R. Liu and I. M. Golovicher, “Energy-efficient operation of rail vehicles,” *Transportation Research Part A: Policy and Practice*, vol. 37, no. 10, pp. 917–931, Oct. 2003.

[17] Y. Wang, B. De Schutter, T. van den Boom, and B. Ning, “Optimal trajectory planning for trains – A pseudospectral method and a mixed integer linear programming approach,” *Transportation Research Part C*, vol. 29, pp. 97–114, Apr. 2013.

[18] J. Pachl, *Railway Operation and Control - Second Edition*. Centralia, USA: Gorham Printing, 2009.

[19] H. P. Williams, *Model Building in Mathematical Programming (4th ed.)*. Chichester, England: John Wiley & Sons, 1999.

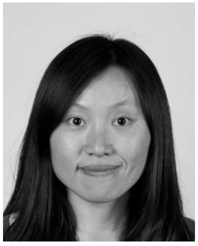
[20] B. Zhang and Y. Lu, “Research on dwelling time modeling of urban rail transit,” *Traffic & Transportation*, vol. 27, no. 12, pp. 48–52, Dec. 2011.

[21] P. Vansteenwegen and D. V. Oudheusden, “Decreasing the passenger waiting time for an intercity rail network,” *Transportation Research Part B*, vol. 41, no. 4, pp. 478–492, May 2007.

[22] S. Wong and C. Tong, “Estimation of time-dependent origindestination matrices for transit networks,” *Transportation Research Part B*, vol. 32, no. 1, pp. 35–48, Jan. 1998.

[23] M. Khan and X. Zhou, “Stochastic optimization model and solution algorithm for robust double-track train-timetabling problem,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 81–89, Mar. 2010.

[24] M. Shafia, M. Aghaee, S. Sadjadi, and A. Jamili, “Robust train timetabling problem: Mathematical model and branch and bound algorithm,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 307–317, Mar. 2012.



Yihui Wang received the B.Sc. degree in control engineering from Beijing Jiaotong University, Beijing, China, in 2007. She is currently working toward the Ph.D. degree in hybrid control of railway systems, with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands. Her research interests include train optimal control, train scheduling, and hybrid systems control.

Bart De Schutter (M’08–SM’10) received the M.Sc. degree in electrotechnical and mechanical engineering and the Ph.D. degree in applied sciences (summa cum laude with congratulations of the examination jury) from the University of Leuven, Leuven, Belgium, in 1991 and 1996, respectively. He is currently a Full Professor with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands. His research interests include the control of intelligent transportation systems, hybrid system control, and multiagent systems.

Bart De Schutter is an Associate Editor for Automatica and for the IEEE Transactions on Intelligent Transportation Systems.

Ton van den Boom received the MSc and PhD degrees in Electrical Engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1988 and 1993, respectively. Currently he is an Associate Professor at Delft University of Technology, The Netherlands. His research interests are in the areas of linear and nonlinear model predictive control, control and identification of discrete event systems and hybrid systems with applications in railway, robotics, and printers.

Bin Ning (M'94) received the B.S. degree from Northern Jiaotong University (now Beijing Jiaotong University), Beijing, China, in 1982, where he received both the M.S. and Ph.D. degrees afterwards. He is currently a Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, as well as the President of Beijing Jiaotong University. His research interests include intelligent transportation systems, communication-based train control, rail transport systems, system fault-tolerant design, fault diagnosis, system reliability, and safety studies.

Prof. Ning is a Fellow of the Institution of Railway Signal Engineers and IET, a Senior Member of the China Railway Society, and a member of the Western Returned Scholars Association. He is the Chair of the Technical Committee on Railroad Systems and Applications of the IEEE Intelligent Transportation Systems Society. Bin Ning is also an Associate Editor for the IEEE Transactions on Intelligent Transportation Systems.

Tao Tang received the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 1991. He is currently a Professor and the Director of the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. He has participated in several major projects on high-speed trains funded by the Chinese government. He is one of the leading scientists in the field with extensive experience in high-speed train control, safety, and energy saving.