Technical report 13-025

Real-time scheduling for single lines in urban rail transit systems

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

If you want to cite this report, please use the following reference instead:

*This report can also be downloaded via [http://pub.deschutter.info/abs/13_025.html](http://pub.deschutter.info/abs/13_025.html)
Real-Time Scheduling for Single Lines in Urban Rail Transit Systems

Yihui Wang, Bart De Schutter, Ton van den Boom  
Delft Center for Systems and Control,  
Delft University of Technology, Delft, The Netherlands  
Email: {yihui.wang,b.deschutter,a.j.j.vandenboom}@tudelft.nl

Bin Ning, Tao Tang  
State Key Laboratory of Rail Traffic Control and Safety,  
Beijing Jiaotong University, Beijing, P.R. China  
Email: {bning,ttang}@bjtu.edu.cn

Abstract—The real-time scheduling problem for urban rail transit systems is considered with the aim of minimizing the total passenger travel time, i.e. the sum of passenger waiting time at stations and passenger on-board time. The operation of trains and passenger demand characteristics are formulated in the real-time scheduling model. The minimum headway constraints are also taken into account to ensure the running safety of trains in urban rail transit. The resulting real-time scheduling problem is a nonlinear non-convex programming problem, which can be solved using state-of-the-art algorithms to obtain the optimal departure times, running times, and dwell times of trains. A case study based on the data of Beijing Yizhuang line is used to demonstrate the performance of the proposed approach.

I. INTRODUCTION

Urban rail transit plays a key role in the public transportation of a city. More specifically, a safe, fast, energy efficient, and comfortable railway system is important for the sustainability of the overall transportation system. With the increasing passenger demand for urban rail transit systems, such as subway systems, the frequency of train operations is very high, especially in these large cities like Beijing, Shanghai, New York, where a train arrives every 3 or 5 minutes. The schedule of trains has a significant effect on the passenger waiting time and the passenger on-board time.

Train scheduling is one of the most challenging problems in railway planning, and it has been studied for decades via different techniques [1], such as the linear programming [2], [3], integer or nonlinear programming [4], [5], [6], [7], and graph theory [8]. In these papers, the available resources, e.g. the single tracks and the crossings, are shared by trains with different origins and destinations. Thus, the trains need to have different times in the train scheduling problem in these papers. However, the lines in urban rail transit usually have double tracks, where each track is used for one direction of train operation. Train overtaking and crossing is normally not allowed during the operations. On the other hand, the passenger demand of urban rail transit vary from station to station during everyday’s operation, which is usually not considered directly in the train scheduling problem in the papers mentioned before. Therefore, the train scheduling problem of urban rail transit is different from that of the interurban rail transit systems. In this paper, we will consider real-time scheduling for urban rail transit.

In the literature, there are many researchers studying the rescheduling of urban transit systems. The rescheduling strategies can be grouped into five general categories [9]: holding, zone scheduling, short turning, deadheading, and dynamic stop-skipping. Holding is used to regulate the headways by holding an early-arriving train, or a train with a relatively short leading headway [10]. In zone scheduling [11], the whole line is divided into several zones, where the trains stop at all stations within a single zone and then run without stopping to the terminal station. The required number of trains and drivers and passenger travel times may be reduced by the zone scheduling, where the zones are defined based on the passenger flows. There are short-turn and full-length trips operating on the line in the short turning strategy [12], [13], where the short-turn trips serve only the zone with high demands and the full-length trips run the whole line. The deadheading strategy involves some trains to run empty through a number of stations at the beginning of their trips to reduce the headways at later stations [9], [14]. The dynamic stop-skipping strategy is frequently used in lines with high demands, which allows those trains that are late and behind the schedule to skip certain low-demand stations and increase the running speed.

All the strategies mentioned above are based on a fixed schedule or a fixed service headway. However, the distance between two stations in urban rail transit is generally less than 3 km and the running time is then less than 3 minutes. As a consequence, if a delay occurs at the platform due to too many passengers boarding the train, then it is not easy to get back to the preplanned timetable because of the short running time. So as different from the state-of-the-art methods in urban rail transit, we propose a real-time scheduling approach to minimize the passenger waiting time and the passenger traveling time, where a predefined timetable or service headway is not needed and the schedule of trains can be optimized in a receding horizon way based on the passenger flows at stations.

The rest of this paper is structured as follows. Section II formulates the model of train movement, the passenger demand characteristics, and the passenger/vehicle interaction. Section III describes the objective function and constraints of the real-time scheduling problem. Section IV proposes to solve the resulting nonlinear non-convex programming problem using sequential quadratic programming approach. Section V illustrates the real-time optimization of the schedule of trains via a case study. We conclude with a short discussion in Section VI.
II. MODEL FORMULATION

In this paper, we consider one direction of a urban transit line consisting of J stations as shown in Figure 1. Station 1 is the origin station and station J is the final station. We assume:

A1. Station j for j \{2,3,\ldots,J-1\} can only accommodate one train at a time and no overtaking can occur at any point in the subway line.

Assumption A1 generally holds for most urban transit systems, which are usually operated in first-in first-out order from station 1 to J. In addition, the order of the running trains is denoted that vehicle \(i\) always precedes train \(i+1\) for \(i \in \{1,2,\ldots,J-1\}\).

A. The model of train movement

In the literature on train scheduling, the detailed dynamics of trains is usually ignored. The movement of trains is described by running times, departure times, and arrival times. In [15], it is assumed that each train runs at a fixed speed except when approaching (or departing from) a station where extra time is required for deceleration (or acceleration). So the running time in [15] is considered as a constant. In this paper, the running time is considered in a certain interval, i.e.

\[
    r_{i,j} \in [r_{i,j,\text{min}}, r_{i,j,\text{max}}],
\]

(1)

where \(r_{i,j}\) is the running time of train \(i\) from station \(j\) to station \(j+1\) as shown in Figure 2, and \(r_{i,j,\text{min}}\) and \(r_{i,j,\text{max}}\) are the minimal and maximal running time for train \(i\) traversing segment \(j\), respectively. The minimal and maximum running time can be calculated based on the detailed train dynamic model, the speed limits, and the grade profiles along the line. Furthermore, the optimal running time will be obtained by solving the real-time scheduling problem.

The departure time \(d_{i,j}\) of train \(i\) at station \(j\) is equal to the sum of the arrival time \(a_{i,j}\) and the dwell time \(\tau_{i,j}\) of train \(i\) at station \(j\), i.e.

\[
    d_{i,j} = a_{i,j} + \tau_{i,j}.
\]

(2)

The dwell time is usually considered as a constant. However, in practice it is influenced by the number of passengers boarding and alighting the train, as will be explained in Section II-C. The arrival time \(a_{i,j+1}\) of train \(i\) at station \(j+1\) equals the sum of the departure time at station \(j\) and the running time on segment \(j\) for train \(i\), i.e.

\[
    a_{i,j+1} = d_{i,j} + r_{i,j}.
\]

(3)

The arrival times and departure times must satisfy the minimum headway constraints caused by the fixed or moving block signaling systems. The minimum headway is the minimum time interval between two successive trains such that they can enter and depart from a station safely [16], i.e.

\[
    a_{i,j} - d_{i-1,j} \geq h_0,
\]

(4)

with \(h_0\) the minimum headway between two trains.

B. Passenger demand characteristic

The passenger arrival rate at station \(j\) for \(j \in \{1,2,\ldots,J\}\) is denoted by \(\lambda_j\). We assume:

A2. Passengers arrive uniformly at a constant rate \(\lambda_j\) at station \(j\).

This assumption is consistent with observed passengers arrivals for short headway (less than 10 minutes) services [17]. An estimate of these passenger arrival rates at stations can be obtained by analyzing historical data of the passenger flow. In addition, we make the following assumption:

A3. The number of passengers waiting at station and the number of passengers on-board immediately after a the departures of trains are real numbers.

Since the number of passengers are large, the error made by this assumption is small. Furthermore, this assumption simplifies the optimization of the train scheduling problem later on.

The number of passengers still remaining at station \(j\) immediately after the departure of train \(i-1\) defined as \(w_{i-1,j}\). The number of passengers who want to get on train \(i\) at station \(j\) can then be formulated as

\[
    w_{i-1,j} + \lambda_j(d_{i,j} - d_{i-1,j}),
\]

where \(\lambda_j(d_{i,j} - d_{i-1,j})\) is number of passengers newly arrived at station \(j\) during the departure of train \(i-1\) and the departure of train \(i\). In addition, we assume:

A4. The number of passengers alighting at station \(j\) for \(j \in \{1,2,\ldots,J\}\) is a fixed proportion \(\rho_j\) of its arrival load.

The passenger alighting proportions can also be estimated using the historical data of the passenger flow.

By defining the number of passengers on train \(i\) immediately after its departure at station \(j-1\) as \(n_{i,j-1}\), the remaining capacity of train \(i\) at station \(j\) immediately after the alighting process of passengers is

\[
    C_{i,\text{max}} - n_{i,j-1}(1 - \rho_j),
\]

where \(C_{i,\text{max}}\) is the maximum capacity of train \(i\) and \(n_{i,j-1}(1 - \rho_j)\) is the number of passengers remaining on train \(i\) immediately after all the passengers that wanted to leave the train have gotten off.
The number of passengers boarding train \( i \) at station \( j \) is equal to the minimum of the remaining capacity and the number of waiting passengers, i.e.

\[
\min \left( C_{i,\text{max}} - n_{i,j-1}(1 - \rho_j), w_{i-1,j} + \lambda_j(d_{i,j} - d_{i-1,j}) \right). \tag{5}
\]

The number of passengers at station \( j \) immediately after the departure of train \( i \), i.e. the passengers who cannot get on train \( i \), is then given by

\[
w_{i,j} = w_{i-1,j} + \lambda_j(d_{i,j} - d_{i-1,j}) - \min \left( C_{i,\text{max}} - n_{i,j-1}(1 - \rho_j), w_{i-1,j} + \lambda_j(d_{i,j} - d_{i-1,j}) \right),
\]

which can be rewritten as

\[
w_{i,j} = \max \left( w_{i-1,j} + \lambda_j(d_{i,j} - d_{i-1,j}) - (C_{i,\text{max}} - n_{i,j-1}(1 - \rho_j)), 0 \right). \tag{6}
\]

In addition, the number of passengers on train \( i \) when it departs from station \( j \) is equal to the sum of the passengers arriving but not getting off at station \( j \) and the passengers boarding on train \( i \) at station \( j \), which can be formulated as

\[
n_{i,j} = n_{i,j-1}(1 - \rho_j) + \min \left( C_{i,\text{max}} - n_{i,j-1}(1 - \rho_j), w_{i-1,j} + \lambda_j(d_{i,j} - d_{i-1,j}) \right). \tag{8}
\]

We can rewrite (8) as

\[
n_{i,j} = \min \left( C_{i,\text{max}}, n_{i,j-1}(1 - \rho_j) + w_{i-1,j} + \lambda_j(d_{i,j} - d_{i-1,j}) \right). \tag{9}
\]

### C. Passenger/vehicle interaction

As mentioned before, the dwell time is influenced by the number of alighting and boarding passengers, which can be described as a linear function of the alighting and boarding passengers [18]. The minimal dwell time can be defined as

\[
\tau_{i,j,\text{min}} = \alpha_1 + \alpha_2 d_{i,j-1} \rho_j + \alpha_3 d_{j-1} \min \left( C_{i,\text{max}} - n_{i,j-1}(1 - \rho_j), w_{i-1,j} + \lambda_j(d_{i,j} - d_{i-1,j}) \right), \tag{10}
\]

where \( \alpha_1, \alpha_2, \alpha_3 \) are coefficients that can be estimated based on historical data. The dwell time \( \tau_{i,j} \) should be larger than the minimal dwell time \( \tau_{i,j,\text{min}} \) such that the passengers can get on and get off the train. However, it should be less than a maximum dwell time \( \tau_{i,j,\text{max}} \) to ensure the passengers do not complain.

### III. THE REAL-TIME SCHEDULING PROBLEM

Based on the model of train movement and the passenger demand characteristic, we now consider the real-time scheduling problem.

In the real-time scheduling problem, the total travel time of all passengers is minimized, which is the sum of the passenger waiting time and the passenger in-vehicle time. The passenger waiting time \( t_{\text{wait},i,j} \) at station \( j \) for train \( i \) includes the waiting time of both passengers left by the previous train \( i-1 \) and the newly arrived passengers, and it can be calculated by

\[
t_{\text{wait},i,j} = w_{i-1,j}(d_{i,j} - d_{i-1,j}) + \frac{1}{2} \lambda_j(d_{i,j} - d_{i-1,j})^2, \tag{11}
\]

where the first term represents the waiting time of the passengers left by train \( i-1 \) at station \( j \), and the second term represents the waiting time of uniformly arriving passengers between the departures of train \( i-1 \) and train \( i \). The passenger in-vehicle time for train \( i \) running from station \( j \) to \( j+1 \) includes the running time for all passengers on train \( i \) after its departure from station \( j \) and the waiting time of the passengers who do not get off the train at station \( j+1 \), which can be formulated as

\[
t_{\text{in-vehicle},i,j} = n_{i,j} \tau_{i,j} + n_{i,j}(1 - \rho_{j+1}) \tau_{i,j+1}. \tag{12}
\]

The total passenger travel time for all \( I \) trains can then be formulated as

\[
t_{\text{total}} = \sum_{i=1}^{I} \sum_{j=1}^{I-1} (t_{\text{wait},i,j} + t_{\text{in-vehicle},i,j}). \tag{13}
\]

The constraints of the real-time scheduling problem consist of the running time constraints, dwell time constraints, headway constraints, capacity of trains, and the equalities mentioned in Section II.

### IV. SOLUTION APPROACH

The variables of the scheduling problem are the departure times \( d_{i,j} \), the running times \( r_{i,j} \), the dwell times \( \tau_{i,j} \), the passengers waiting at stations \( w_{i,j} \), and the passengers on-board the trains \( n_{i,j} \) for \( i = \{1, 2, \ldots, I\} \) and \( j = \{1, 2, \ldots, J - 1\} \). The other variables are auxiliary variables, such as the passenger waiting times \( t_{\text{wait},i,j} \), the passenger on-board times \( t_{\text{in-vehicle},i,j} \). After the elimination of the auxiliary variables, the real-time scheduling problem formulated in the previous section becomes a nonlinear non-convex programming problem with nonlinear, non-convex constraints (due to the min function).

Hence, multi-start local optimization (e.g. sequential quadratic programming (SQP), genetic algorithm, simulated annealing, active-set algorithm) [19, Section 5.3] can be used to solve the problem. In this paper, we will apply the SQP algorithm to the real-time scheduling problem.

### V. CASE STUDY

In order to demonstrate the performance of the real-time scheduling approach proposed in this paper, the line data and train characteristics of the Yizhuang subway line in Beijing are used as a test case study. There are 14 stations in the Yizhuang line and the total length is 22.773 km. The speed limit for the whole line is 80 km/h, i.e. 22.2 m/s. The detailed information of the Yizhuang line and the parameters used in the case study are listed in Table I. Note that the passenger arrival rate at the final station Yizhuang is 0 passenger/s since we only consider one direction of the line. In addition, the proportion of the alighting passengers is also listed in Table I. There are no passengers alighting the train at the origin and all passengers on the train should get off at the final station.

Furthermore, in this paper the minimum running time between these stations is calculated by taking a fixed acceleration and a fixed deceleration, which are both chosen as 0.8 m/s\(^2\) in absolute value. In addition, the train is assumed to run at the maximum speed \( v_{\text{max}} = 22.2 \) m/s between the acceleration phase and the deceleration phase. Therefore, the minimum running time between station \( j \) and station \( j+1 \) with an intersection distance \( s_j \) can be calculated as

\[
r_{i,j,\text{min}} = \frac{s_j}{v_{\text{max}}} + t_{\text{acc-dec}}, \tag{14}
\]
Fig. 3. The layout of the Yizhuang subway line

**TABLE I. THE INFORMATION OF THE YIZHUANG SUBWAY LINE AND OTHER PARAMETERS**

<table>
<thead>
<tr>
<th>Station number</th>
<th>Station name</th>
<th>Distance to next station [m]</th>
<th>Passenger arrival rate [passenger/s]</th>
<th>Passenger alighting proportion</th>
<th>Minimum running time [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Songjiazhuang</td>
<td>1332</td>
<td>2</td>
<td>0</td>
<td>87.700</td>
</tr>
<tr>
<td>2</td>
<td>Xiaocun</td>
<td>1326</td>
<td>2</td>
<td>0.1</td>
<td>85.628</td>
</tr>
<tr>
<td>3</td>
<td>Xiaohongmen</td>
<td>2086</td>
<td>2</td>
<td>0.3</td>
<td>121.664</td>
</tr>
<tr>
<td>4</td>
<td>Jiugong</td>
<td>2265</td>
<td>4</td>
<td>0.5</td>
<td>129.727</td>
</tr>
<tr>
<td>5</td>
<td>Yizhuangqiao</td>
<td>2331</td>
<td>4</td>
<td>0.3</td>
<td>132.700</td>
</tr>
<tr>
<td>6</td>
<td>Wenhuyuan</td>
<td>1354</td>
<td>4</td>
<td>0.4</td>
<td>88.691</td>
</tr>
<tr>
<td>7</td>
<td>Wanyuan</td>
<td>1280</td>
<td>3</td>
<td>0.3</td>
<td>85.358</td>
</tr>
<tr>
<td>8</td>
<td>Rongjing</td>
<td>1544</td>
<td>2</td>
<td>0.4</td>
<td>97.250</td>
</tr>
<tr>
<td>9</td>
<td>Rongchang</td>
<td>992</td>
<td>3</td>
<td>0.2</td>
<td>72.385</td>
</tr>
<tr>
<td>10</td>
<td>Tongjinan</td>
<td>1975</td>
<td>4</td>
<td>0.4</td>
<td>116.664</td>
</tr>
<tr>
<td>11</td>
<td>Jinhai</td>
<td>2389</td>
<td>3</td>
<td>0.4</td>
<td>134.412</td>
</tr>
<tr>
<td>12</td>
<td>Ciquan</td>
<td>1339</td>
<td>2</td>
<td>0.7</td>
<td>88.466</td>
</tr>
<tr>
<td>13</td>
<td>Ciqu</td>
<td>2610</td>
<td>1</td>
<td>0.6</td>
<td>145.268</td>
</tr>
<tr>
<td>14</td>
<td>Yizhuang</td>
<td>-</td>
<td>0</td>
<td>1.0</td>
<td>120.0</td>
</tr>
</tbody>
</table>

Table II. THE PARAMETERS OF THE MINIMAL DWELL TIME

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_{i,j} )</td>
<td>0.002</td>
</tr>
<tr>
<td>( \beta_{i} )</td>
<td>0.0503</td>
</tr>
<tr>
<td>( \gamma_{i} )</td>
<td>0.3466</td>
</tr>
</tbody>
</table>

where \( s_{\text{acc-dec}} \) and \( t_{\text{acc-dec}} \) are the running distance and the running time of the acceleration and deceleration phase, respectively. The computed minimum running times are also shown in Table I. The maximum running time is assumed as

\[
r_{i,j,\text{max}} = \zeta r_{i,j,\text{min}},
\]

where \( \zeta \) is larger than 1. We have chosen \( \zeta \) as 1.5 in order to ensure the passengers do not complain the train is too slow.

The dwell time at stations has a significant effect on the real-time scheduling problem for urban rail transit. Based on the dwell time research about Beijing subway stations [20], the value of the dwell time coefficients are chosen as shown in Table II. The minimal dwell time can then be calculated by (10). The maximal dwell time is chosen as 150 s. The capacity of each train is 1468 passengers according to the train characteristics in the Yizhuang line. In addition, the minimum headway \( h_{0} \) between two successive trains is 90 s.

In this case study, we consider the real-time scheduling problem for 7 trains, i.e. train \( i \in \{1, 2, \ldots, 7\} \). In addition, it is assumed that train 0 precedes train 1. The schedule of train 0 shown as the red line in Figure 4 is already given and fixed, where it arrives at the platform of station 1 at 0 s, runs with the minimum running time, and stops at each station for 120 s. Furthermore, we assume that there are no passengers left by train 0, i.e. \( w_{0,j} = 0 \) for all \( j \in \{1, 2, \ldots, 13\} \). We use the SNOPT solver via the Tomlab interface to Matlab.

**Fig. 4. The computed schedule for trains**

for solving the nonlinear non-convex scheduling problem [21]. By solving the real-time scheduling problem for the following 7 trains using an SQP algorithm [22], the total passenger travel time for the passengers traveled by these 7 trains is \( 2.1047 \times 10^{7} \) s. The resulting schedules of these trains are shown in Figure 4. The departure times of these 8 trains at station 1 are 120, 360, 600, 840, 961.2, 1065.7, 1170.3, and 1274.8, respectively. The headways between these trains are then 240 s, 240 s, 240 s, 121.2 s, 104.6 s, 104.6 s, and 104.6 s, respectively, which are shown in Figure 5. The departure headway between train 0 and train 1 is quite large, and it is equal to the sum of the minimum headway 90 s and the maximum dwell time 150 s. This is because of the schedule of train 0, which stops at each stations with dwell time 120 s. Therefore, in order to satisfy the headway constraints at all stations, the departure headway at the station 1 must be much larger than the minimum headway 90 s. As we can observe from Figure 5, most of the headways at stations are around 153 s, especially for the trains and stations with a higher index. This is because passenger waiting times are closely related to vehicle headways and evenly distributed headways usually result in the least passenger waiting times when the passenger arrival is a uniform process and the arrival rate is constant [15]. The calculated running times and dwell times of different trains at different stations are shown in Figure 6 and Figure 7.

The number of passengers waiting at each station and the number of passengers on each train immediately after the train’s departure are shown in Figure 8 and Figure 9. As we can see from Figure 8, there is no passenger left by these 7 trains at station 1, 2, 3, 4, 8, 11, 12, and 13. There are 11 passengers left by train 2 at station 5. The maximal number of waiting passengers is left by train 6 at station 7 and the number is 225 passengers. The number of waiting passengers at station 6 increases for the first 6 trains and it decreases after the departure of train 7. If we let more trains run, then the number of waiting passengers will decrease further like that of station 6. The different passenger arrival rates and passenger alighting proportions at different stations have a significant effect on the number of waiting passengers and on-board passengers. As we can see from Table I, the passenger arrival rates at station 4, 5, and 6 are equal to 4 passenger/s. Therefore,
the number of passengers on-board increases quickly when trains passing through these stations. On the other hand, the passenger alighting proportions at station 5 and 6 are 0.3 and 0.4, respectively. Hence, there are not too many passengers getting off the trains at these two stations. That is the reason why there are many passengers left at station 6. The passenger arrival rate at station 7 is 3 passenger/s, which is less than 4 passenger/s, but the passenger alighting proportion is 0.3. In addition, the number of passengers waiting at station 7 is also affected by the situation in station 6. Therefore, there are also many passengers waiting at station 7. The passengers on-board immediately after the departure of each train is less than or equal to 1468 since the capacity of each train is 1468 passengers. The number of passengers on-board first increases and then reaches the capacity. At the end of the line the number of passengers is getting less since the passenger arrival rate is lower and the passenger alighting proportion is higher.

**VI. CONCLUSIONS**

In the current paper, we have considered the real-time scheduling problem for urban rail transit. The model of train movement has been described by the departure times, the running times, and the dwell times at stations. The minimum
headway constraints were also taken into account to ensure the running safety of all trains. The passenger arrival rates and alighting proportions have been used to represent the passenger demand characteristics. The resulting nonlinear non-convex scheduling problem minimizing the total passenger travel times has been solved by the sequential quadratic programming algorithm. The case study shows that the optimal headways between these trains are not a constant which is usually used in the planning phase of urban rail transit, but the optimal headways are affected by the passenger demands at stations and the current situation in the urban transit (the schedule of the initial train, the initial waiting passengers at stations, etc.).

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