Optimal Train Service Design in Urban Rail Transit Line with Considerations of Short-Turn Service and Train Size

Zhengyang Li a,b, Jun Zhao a,b,1, Qiyuan Peng a,b

a School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, Sichuan 611756, China
b National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu, Sichuan 611756, China

1 Corresponding author, E-mail: junzhao@swjtu.edu.cn

Abstract
The train service scheme of an urban rail transit line specifies information such as the total number of train services operated in the line, and the associated turn-back stations, train size and frequency of each service. A reasonable train service scheme can provide satisfactory services for passengers and reduce the operational cost for operators. This paper focuses on the optimal train service design problem in an urban transit line, where both the short-turn services and the train size of each service are considered. A service network based on a given pool of candidate train services with provided turn-back stations is constructed. The optimal strategy is used to assign passenger flows on the service network so as to describe the service choice behaviour of passengers between different train services. Considering many operational and capacity constraints, a mixed integer nonlinear programming model minimizing the sum of the operators’ cost and passengers’ waiting time cost is developed to identify train services from the service pool and determine the train size and frequency of each chosen service. The nonlinear model is transformed into a linear one, and two simplification methods named service network simplification and OD pair aggregation are proposed to improve further the computational efficiency of the model. Finally, realistic instances from Chongqing Rapid Rail Transit Line 26 in China are used to test the proposed approaches. The results show that our approach can effectively reduce the operators’ cost and the passengers’ waiting time cost compared with the empirical method frequently used in practice.

Keywords
Urban rail transit, train service design, short-turn service, train size, service network, optimal strategy

1 Introduction

The train service scheme plays an important role in the operation of urban rail transit. It contains information such as the total number of train services in a line, the turn-back stations, train size and frequency of each designed service. The train service scheme affects the waiting time and transfer time of passengers, and it also determines the number of rolling stocks and crews that required to operate an urban rail transit line. A reasonable train service scheme can match the transport capacity with the passenger demand of a line, leading to a reduction of the passengers’ waiting time and the operation cost of the line.

At present, using a single full-length train service and determining the associated train size and train frequency according to the maximum passenger load of sections is still a
frequently used method to obtain the train service scheme in an urban rail transit line. In this way, the train frequency is the same in every section of a line. However, this empirical method does not take the possible unbalanced spatial distribution of passenger demand in urban rail transit lines into account, resulting in waste of transport capacity in some sections and congestion in other sections. A commonly used strategy to overcome this problem is to insert short-turn services into the train service scheme. Short-turn services enable different sections to have different train frequencies, where the transport capacity can match the passenger demand better. However, the setting of short-turn services will cause some passengers to transfer, and passengers’ service choice behaviour is complex in the case of multiple train services, which complicates the train service scheme design. Therefore, it is necessary to analyse the impact of train services on passengers and operators, and optimize the train service scheme with multiple services to improve the service quality and reduce the operation cost of urban rail transit lines.

This paper tries to optimize the train service scheme in an urban rail transit line where multiple train services consisting of either full-length or short-turn ones can be operated, and the train size and frequency of each service have to be determined. In order to accurately describe the service choice behaviour of passengers between different train services, a service network based on a given pool of candidate train services with specified turn-back stations is constructed, and the optimal strategy proposed by Spiess (1989) is used to assign passenger flows on the service network. A mixed integer nonlinear programming model minimizing the sum of the operators’ cost and passengers’ waiting time cost is developed to identify train services from the service pool and determine the train size and frequency of each chosen service. Then, the nonlinear model is transformed into a linear one which is further simplified by two methods, enabling the linear model to be solved quickly by commercial optimization solver CPLEX. Finally, instances based on the Chongqing Rapid Rail Transit Line 26 in China are constructed to test the proposed approaches.

The remainder of this paper is structured as follows. Section 2 gives an overview of the related literature. In Section 3, we present the problem description and assumptions. In Section 4, a mixed integer nonlinear programming model is formulated to represent the train service design problem in an urban rail transit line. Section 5 presents a linearization method and two simplification techniques to obtain a simplified mixed integer linear programming model. Section 6 provides our computational experiments on Chongqing Rapid Rail Transit Line 26 in China. Conclusions and future research works are discussed in Section 7.

2 Literature Review

To the best of our knowledge, our problem has not been completely investigated in the literature. Related works mainly focus on the timetable design with short-turn services or multiple vehicle sizes, while few works study the service design in public transport systems especially in urban rail transit lines.

There are few works focusing on obtaining the timetable of public transport systems with short-turn services or multiple vehicle sizes. Furth (1987) considered that the schedule coordination between the full-length trip and short-turn trip is necessary, and proposed an offset schedule algorithm to minimize the bus fleet size and to save the operation cost. Ceder (1989) proposed a two-stage optimization method to obtain the location of turn-back stations and the bus fleet size. Zhang (2018) developed a mixed integer linear programming model to optimize the timetable of an urban rail transit line with short-turn strategy and multiple depots. With the consideration of multiple vehicle sizes, Ceder and Hassold (2011) proposed a multi-objective methodology to create even load and even headway bus
timetables by operating different bus sizes. Hassold and Ceder (2012) presented approaches to use multiple vehicle sizes to improve the matching between bus timetable and passenger demand. Chen (2019) considered the design of headway and vehicle capacity simultaneously with the usage of modular vehicles, aiming to better match the dynamic passenger demand with transit services.

Other works mainly studied the transit service design problem with short-turn services in either bus corridors or urban rail transit lines. Delle Site and Filippi (1998) focused on the short-turn strategies with different bus sizes within multiple operating periods, and developed a net benefit maximization model for optimizing the bus sizes, service frequencies and fares. Tirachini (2011) developed a model to optimize the short-turn trip in a single period by analytical expressions, aiming at increasing the bus frequency in congestion sections. Cancela (2015) considered the interests of both operators and users, and used the optimal strategy proposed by Spiess (1989) to develop a mixed integer linear model to solve the bus routes design problem. Ji (2016) used a Markov model to describe the seats searching behaviour of passengers during their trip, and proposed a model to optimize the schedule coordination between full-length and short-turn bus services. Sun (2016) relaxed the assumption that a full-length service must be operated, and proposed a flexible short-turn service design model to minimize the operators’ cost and passengers’ waiting time in subways. Yang (2017) developed a bilevel model to design the short-turn strategy on a bus route. Ding (2018) relaxed the constraints of turn-back stations in metro systems, and proposed a nonlinear programming model to design the short-turn services.

A review of existing studies can be summarized as follows:

1. Most of the previous studies focus on the timetable optimization in both bus and rail systems, while few works focus on the service design problem in transit lines, especially in urban rail transit lines. Compared with the bus systems, urban rail transit lines have more restrictions on setting train services, especially the capacity limitation of turn-back stations.

2. Due to the difficulty of depicting the services setting on either bus or rail lines and the complexity of analysing the service choice behaviour of passengers in the case of multiple services. Almost all the former works follow an assumption that a full-length service must exist, and the problem is to obtain the optimal parameters of setting one short-turn service. But, when additional services are considered, many existing models become intractable. At present, there are already some urban rail transit lines which have more than two train services, such as Shanghai Metro Line 2 (3 services) and Chongqing Rail Transit Line 3 (3 services) in China. Therefore, it is necessary to continue to study the optimization method for multiple train services design in urban rail transit lines.

3. Under the assumption that a full-length service exists, most of the initial works assume that passengers can always take a direct service to their destination when designing the service scheme. Few studies have been done to describe passengers’ service choice behaviour during their trip, especially in the case of multiple services.

4. Relative works mainly consider the usage of multiple vehicle sizes in timetable optimization. But in the aspect of service design, most of the existing studies only design short-turn services whose vehicle size is pre-determined. Actually, the train size interacted with the frequency is an important parameter of services which affects the capacity and operation cost of urban rail transit lines. Joint design of the train size and frequency of services could lead to better solutions.
3 Problem Description and Assumptions

3.1 Problem Description

Without loss of generality, we consider an urban rail transit line with several turn-back stations, multiple services and multiple train sizes. Taking an urban rail transit line with 7 stations and 3 turn-back stations shown in Figure 1 as an example. The stations are denoted as \( v_1 \) to \( v_7 \) in the upward direction, while the sections are denoted as \( e_1 \) to \( e_6 \) in the upward direction. Stations \( v_1, v_4 \) and \( v_7 \) are turn-back stations on the line that can reverse the running direction of trains. Restricted by the layout of the turn-back tracks, stations \( v_1, v_7 \) can only reverse the running direction for trains from one direction (\( v_1 \) can only switch the running direction of trains from downward to upward, and \( v_7 \) can only switch the running direction of trains from upward to downward), while station \( v_4 \) can reverse the running direction for trains from both directions.

\[
\begin{array}{cccccccc}
  & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
 v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & \\
\end{array}
\]

Figure 1: Sketch map of an urban rail transit line

According to the given location of turn-back stations on the line and the feasible reversing directions of each turn-back station, a pool of candidate train services can be generated. The line shown in Figure 1 can operates 3 train services shown in Figure 2. Service 1 has stations \( v_1 \) and \( v_7 \) as its turn-back stations. Service 2 has stations \( v_1 \) and \( v_4 \) as its turn-back stations. Service 3 has stations \( v_4 \) and \( v_7 \) as its turn-back stations.

\[
\begin{array}{cccccccc}
  & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & \\
  & & & & & & & & \\
 \end{array}
\]

Figure 2: Illustration of train services

In an urban rail transit line, after the construction of candidate train services, the train service design problem is to determine the train services operated on the line, and the train size and frequency of each selected train service. To design the train service scheme, we need to take into account the operation cost changes between different schemes as well as the impacts of schemes on the travel process of passengers. In addition, a feasible train service scheme must consider several operational and capacity constraints, e.g. coverage of stations and sections, passenger load in sections, capacity of turn-back stations, operational rules of train services, etc. Therefore, the train service design problem of an urban rail transit line is essentially to find the optimal combination of train services, and the train size and frequency of each selected service under the constraints of line condition and passenger demand.

Formally, in an urban rail transit line, given the layout of the line, capacity of turn-back stations, minimal and maximum frequency in sections, passenger load in sections and other operational rules of train services, the train service design problem on an urban rail transit line is to determine the train service set on the line, the train size of each selected service...
and the corresponding train frequency such that all operational and capacity constraints are respected, while both the operators’ cost and passengers’ waiting time cost are minimized.

3.2 Assumptions

To simplify the model formulation, the following assumptions are introduced.

(1) Passengers’ behaviour accords with the principle of optimal strategy, that is, at each station, each passenger boards the first train passing the station, which can transport him/her close to his/her destination station.

(2) At each station, passengers arrive uniformly and trains arrive on timetable.

(3) Trains are used independently in each service. Only one train composition is allowed in each service. Different train compositions can be arranged to different services.

(4) Trains of each service are assumed to stop at each station of the service route.

4 Model Formulation

4.1 Service Network Construction

According to the given layout especially the turn-back stations of an urban rail transit line, all candidate train services that are allowed to be operated on the line can be generated in advance. Based on which, a directed service network is introduced to design the train services and describe the travel process of passengers in the studied line. Taking the line in Figure 1 as an example, as there are 3 candidate train services in the line, a directed service network shown in Figure 3 can be formed.

The node set of the service network consists of two parts, including station nodes 1 to 7 and train service nodes 8 to 37. Station nodes 1 to 7 represent the stations \( v_1 \) to \( v_7 \) in the line, while train service nodes indicate the stations covered by all candidate services. Service 1 covers stations \( v_1 \) to \( v_7 \). The red nodes 8-14 and 23-29 with respect to Service
1 cover station nodes 1 to 7 in parallel. Service 2 covers stations $v_1$ to $v_4$. The green nodes 15-18 and 30-33 with regard to Service 2 cover station nodes 1 to 4 in parallel. Service 3 covers stations $v_4$ to $v_7$. The blue nodes 19-22 and 34-37 corresponding to Service 3 cover station nodes 4 to 7 in parallel. Among the train service nodes, nodes 8 to 22 are in the upward direction, and nodes 23 to 37 are in the downward direction.

The arc set is composed of three parts including the boarding arcs, running arcs and alighting arcs. Boarding arcs connect the station nodes to its corresponding train service nodes, indicating the boarding process of passengers from stations to the trains of each service in each direction. In both directions, running arcs sequentially link the train service nodes of each service, expressing the running process of trains through sections in each direction of each service. Alighting arcs are from the train service nodes to the corresponding station nodes, representing the alighting process of passengers from the trains of each service in each direction to the corresponding stations. Note that in both directions, boarding arcs do not connect the station node to the last node of each service, and alighting arcs do not connect the first node of each service to the corresponding station node. For instance, in the upward direction, there is no a boarding arc between node 4 and node 18, because the trains of Service 2 cannot take the passengers from station $v_4$ to the upward direction. Also, there is no an alighting arc between node 19 and node 4, because station $v_4$ is the origin station of Service 3 to the upward direction and the trains of Service 3 are empty when originating from station $v_4$ due to that no passengers need to alight.

After abstracting the considered urban rail transit line into a directed service network, the travel process (including boarding, in-vehicle running, transferring and alighting) of each passengers OD pair on the line can be expressed by a path from its origin station node to its destination station node in the service network. For example, the journey of passengers in OD pair $(1, 6)$ from station $v_1$ to station $v_6$ on the line in Figure 1 can be represented by a path from station node 1 to station node 6 in the service network of Figure 3 as follows:

1. If passengers take the trains of Service 1 directly from station $v_1$ to station $v_6$, their travel process can be expressed by the path 1-8-9-10-11-12-13-6 in the service network.
2. If passengers firstly take the trains of Service 2 from station $v_1$ to station $v_4$, and then transfer to trains of Service 3 until arriving at station $v_6$, the travel process is represented by the path 1-15-16-17-18-4-19-20-21-6 in the service network.
3. If passengers from station $v_1$ to station $v_6$ sequentially take the trains of Service 2 and Service 1 with a transfer at station $v_4$, their travel process can be indicated by the path 1-15-16-17-18-4-11-12-13-6 in the service network.

In order to analyse the impact of different train service schemes on the service choice behaviour of passengers, not only the travel process of passengers on the line but also the distribution of passenger flows on each arc of the service network should also be determined. Here we adopt the optimal strategy proposed by Spiess (1989) to assign the passengers of each OD pair on the service network. Under the optimal strategy, passengers take the first train which can transport them close to their destination station. Due to that, at any station node, the probability that a passenger chooses a boarding arc originated from this station node to the travel direction of the passenger is the ratio of the frequency of the service corresponding to the boarding arc to the total frequency of all boarding arcs originating from this station node to the travel direction of this passenger. With the optimal strategy, the distribution of each OD pair on the service network is obtained. Thus, the impact of different train service schemes on the trip decisions of passengers can be quantitatively analysed.
4.2 Notation

The notation to be used in the model is provided in Table 1.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V$</td>
<td>Set of stations of an urban rail transit line with index $v$.</td>
</tr>
<tr>
<td>$E$</td>
<td>Set of sections with index $e$.</td>
</tr>
<tr>
<td>$L$</td>
<td>Set of candidate train services with index $l$.</td>
</tr>
<tr>
<td>$T$</td>
<td>Set of optional train sizes with index $t$.</td>
</tr>
<tr>
<td>$D$</td>
<td>Set of running directions with index $d$.</td>
</tr>
<tr>
<td>$F$</td>
<td>Set of optional trains with index $f$.</td>
</tr>
<tr>
<td>$N$</td>
<td>Set of nodes in the service network with index $n$, $N = {N_1, N_2}$.</td>
</tr>
<tr>
<td>$N_1$</td>
<td>Set of station nodes.</td>
</tr>
<tr>
<td>$N_2$</td>
<td>Set of train service nodes.</td>
</tr>
<tr>
<td>$A$</td>
<td>Set of arcs in the service network with index $a$, $A = {A_1, A_2}$.</td>
</tr>
<tr>
<td>$A_1$</td>
<td>Set of boarding arcs, $A_1 = {A_{1d}</td>
</tr>
<tr>
<td>$A_{1d}$</td>
<td>Set of boarding arcs in direction $d$.</td>
</tr>
<tr>
<td>$A_2$</td>
<td>Set of running and alighting arcs.</td>
</tr>
<tr>
<td>$A_{n+}$</td>
<td>Set of outgoing arcs from node $n$.</td>
</tr>
<tr>
<td>$A_{n-}$</td>
<td>Set of incoming arcs to node $n$.</td>
</tr>
<tr>
<td>$\lambda_n$</td>
<td>Corresponding station of node $n$ in the service network.</td>
</tr>
<tr>
<td>$\tau_a$</td>
<td>Time span of the study period, unit: minute.</td>
</tr>
<tr>
<td>$c^t_l$</td>
<td>Fixed cost of a train with size $t$ within the study period.</td>
</tr>
<tr>
<td>$c^t_e$</td>
<td>Operating cost of a train per kilometre with train size $t$.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Passenger’s waiting time cost per hour.</td>
</tr>
<tr>
<td>$m_l$</td>
<td>Round-trip time of train service $l$.</td>
</tr>
<tr>
<td>$g_l$</td>
<td>Round-trip distance of train service $l$.</td>
</tr>
<tr>
<td>$\alpha_{l_0}$</td>
<td>0-1 parameters, if service $l$ covers station $v$, $\alpha_{l_0}=1$, 0 otherwise.</td>
</tr>
<tr>
<td>$\beta_{l_e}$</td>
<td>0-1 parameters, if service $l$ covers section $e$, $\beta_{l_e}=1$, 0 otherwise.</td>
</tr>
<tr>
<td>$\gamma_{ltd}$</td>
<td>0-1 parameters, if trains in direction $d$ of service $l$ turn back at station $v$, $\gamma_{ltd}=1$, 0 otherwise.</td>
</tr>
<tr>
<td>$q_{ij}$</td>
<td>Volume of OD pair $(i, j)$ from station $i$ to station $j$ within the study period.</td>
</tr>
<tr>
<td>$b_{ij}$</td>
<td>Travel direction of OD pair $(i, j)$, $b_{ij} \in D$.</td>
</tr>
<tr>
<td>$p_{ed}$</td>
<td>Passenger load at section $e$ in direction $d$ within the study period, which can be obtained from the passenger OD demand.</td>
</tr>
<tr>
<td>$r_t$</td>
<td>Capacity of a train with size $t$.</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Required surplus of transport capacity in sections, unit: %.</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>Maximum allowable number of train services on the studied line.</td>
</tr>
<tr>
<td>$\phi$</td>
<td>Minimum train frequency of each service.</td>
</tr>
<tr>
<td>$s_v$</td>
<td>Turn-back capacity of station $v$ in direction $d$ within the study period.</td>
</tr>
<tr>
<td>$b_{min}^e$</td>
<td>Minimal train frequency requirement of section $e$ within the study period.</td>
</tr>
<tr>
<td>$b_{max}^e$</td>
<td>Maximal train frequency limitation of section $e$ within the study period.</td>
</tr>
<tr>
<td>$x_{lt}$</td>
<td>0-1 variables, if train size $t$ is adopted on service $l$, $x_{lt}=1$, 0 otherwise.</td>
</tr>
<tr>
<td>$y_{l_0t}$</td>
<td>0-1 variables, if the $f$th train with size $t$ on service $l$ is operated, $y_{l_0t}=1$, 0 otherwise. Thus, the frequency of trains with size $t$ on service $l$ is $\sum_{f \in F} y_{l_0f}$.</td>
</tr>
<tr>
<td>$k_{a}$</td>
<td>Continuous variables, represents the volume (number of passengers) of OD pair $(i, j)$ on arc $a$ of the service network.</td>
</tr>
<tr>
<td>$w_{in}$</td>
<td>Continuous variables, represents the total passenger waiting time of OD pair $(i, j)$ at station node $n$ of the service network.</td>
</tr>
</tbody>
</table>
4.3 Operator’s Cost

The operator’s cost of an urban rail transit line contains the fixed cost and operation cost. Fixed cost refers to the cost of purchasing rolling stocks used in the line. The rolling stocks of an urban rail transit line are not only operated in the study period, and the life cycle of a rolling stock is typically 30 years. Therefore, to define comparable cost items, the fixed cost of operators in purchasing rolling stocks is apportioned by the time span of the study period. Meanwhile, operation cost refers to the total operating cost of all trains running on the line during the study period, which depends on the round-trip distance of all chosen train services. The fixed cost and operation cost of operators are formulated as follows:

\[ Z_1 = \sum_{t \in T} \sum_{l \in L \cap T} \sum_{f \in F_t} \frac{c_{1n} y_{nlf}}{h} \]  

Objective (1) represents the fixed cost of operators. Note that the number of rolling stocks required by each train service equals to the round-trip time of the service multiplied by the frequency of the service. Objective (2) indicates the operation cost of operators.

4.4 Passengers’ Waiting Time Cost

The travel time of passengers in an urban rail transit line is mainly composed of four parts including the access time, origin/transfer waiting time, in-vehicle time and egress time. The train service scheme rarely impacts the access time and egress time of passengers. Meanwhile, trains with different sizes always have the same running time in each section of the line, due to that different train service schemes make no difference to the in-vehicle time of passengers. However, under multiple train services, the origin, destination and frequency of operated services are different, which could lead to different waiting time of passengers. Therefore, we focus on minimizing the waiting time cost of passengers.

In theory, the waiting time of passengers at each station to each direction is the average waiting time of passengers multiplied by the number of passengers at the station to the direction. Following the optimal strategy in Spiess (1989), assume that passengers arrive at stations uniformly and each passenger boards the first train which can transport them close to their destination station, i.e. passengers show no preference for the choice of services. Moreover, in the practical operation of urban rail transit lines, the arrival of the trains regulated by timetable at each station to each direction is deterministic and even. Therefore, the expected average waiting time of passengers at a station node to a direction is half of the combined arrival interval of the services corresponding to all boarding arcs of the station to the direction.

The expected average waiting time of passengers at station node \( n \) to direction \( d \) is:

\[ E(wt) = \frac{h}{2 \sum_{a \in A} \sum_{t \in T} \sum_{f \in F_t} y_{atf}} \]  

Meanwhile, the number of passengers at a station node to a direction is the sum of volume of OD pairs on all boarding arcs of the station to the direction, which can be obtained by the passenger flow distribution on the service network. Thus, the waiting time cost of passengers on the whole line is:
\[ Z_3 = \frac{\sigma}{60} \sum_{\ell \in L} \sum_{f \in F} \sum_{t \in T} \sum_{n \in N} h \sum_{a \in A_{n+} \cap A_{f \ell}} k_{ij}^{ij} \frac{k_{ij}^{ij}}{2 \sum_{a \in A_{n+} \cap A_{f \ell}} \sum_{t \in T} \sum_{f' \in F} Y_{t, f' t, f} } \] (4)

4.5 Basic Model

Based on the above modelling, the original compound objective function to minimize the operators’ cost and passengers’ waiting time cost is as follows:

\[
\min \sum_{\ell \in L} \sum_{f \in F} \sum_{t \in T} \sum_{n \in N} \frac{c_{i}^{i} y_{i f}}{h} + \sum_{\ell \in L} \sum_{f \in F} \sum_{t \in T} \sum_{n \in N} c_{i}^{i} y_{i f} + \frac{\sigma}{60} \sum_{\ell \in L} \sum_{f \in F} \sum_{t \in T} \sum_{n \in N} w_n^{ij} \] (4)

Note that the third part of Objective (4) which represents the waiting time cost of passengers is a nonlinear representation which will cause difficulties in solving the model.

We utilize the optimal strategy proposed in Spiess (1989) to assign passengers in the service network such that a linear representation of the passengers’ waiting time can be obtained. Then, we formulate our problem as the following basic and incomplete model:

\[
\min \sum_{\ell \in L} \sum_{f \in F} \sum_{t \in T} \sum_{n \in N} \frac{c_{i}^{i} y_{i f}}{h} + \sum_{\ell \in L} \sum_{f \in F} \sum_{t \in T} \sum_{n \in N} c_{i}^{i} y_{i f} + \frac{\sigma}{60} \sum_{\ell \in L} \sum_{f \in F} \sum_{t \in T} \sum_{n \in N} w_n^{ij} \] (6)

s.t. \[
\sum_{a \in A_{n+}} k_{ij}^{ij} - \sum_{a \in A_{n-}} k_{ij}^{ij} = \begin{cases} q_{ij}, & \lambda_n = i, n \in N_1, \forall i, j \in V \\ -q_{ij}, & \lambda_n = j, n \in N_1, \forall i, j \in V \\ 0, & n \in N_2, \forall i, j \in V \end{cases} (7)
\]

\[
k_{ij}^{ij} \leq \frac{2}{h} w_n^{ij} \sum_{t \in T} \sum_{f \in F} y_{t, f}, \quad \forall a \in A_{n+} \bigcap A_{1 \beta_{ij}}, \forall n \in N_1, \forall i, j \in V (8)
\]

\[
k_{ij}^{ij} = 0, \quad \forall a \in A_{n+} \bigcap A_{1 \beta_{ij}}, \forall n \in N_1, \forall i, j \in V (9)
\]

Objective function (6) minimizes the operators’ cost and the passengers’ waiting time cost. Constraints (7) are the flow conservation of OD pairs, which assure that each OD pair can be routed from its origin to its destination in the service network. Constraints (8) are the optimal strategy requirements imposed on decision variables $y_{i f}$, $k_{ij}^{ij}$ and $w_n^{ij}$. These constraints mean that for each OD pair $(i, j)$, the number of passengers of this pair $k_{ij}^{ij}$ on each boarding are $\alpha$ originating from each station node $n$ is not greater than the quotient between the waiting time of all passengers of this pair $w_n^{ij}$ at station node $n$ and 1/2 of the arrival interval of the service $h/\sum_{t \in T} \sum_{f \in F} y_{t, f}$ in boarding arc $a$. Constraints (8) are specially proposed to ensure that the assignment of OD pairs satisfy the optimal strategy, and it can help to obtain a linear representation of the waiting time cost of passengers in the objective function. Constraints (9) indicate that passengers do not take trains which would carry them away from their destination. In other words, there are no flows on the boarding arcs to the opposite travel direction of each OD pair. Constraints (10) and (11) are the non-negative constraints of decision variables.

The basic model (6) to (11) mainly conducts the passenger assignment in the service network under the optimal strategy, aiming to obtain the distribution of each OD pair on the
service network and compute the total waiting time cost of passengers. The operator’s fixed
cost and operation cost are computed and other operational and capacity requirements are
respected by incorporating the following additional constraints.

4.6 Additional Constraints

4.6.1 Covering of stations and sections
The train services operated on an urban rail transit line need to cover each station and each
section of the line. These requirements are as follows:

\[
\sum_{t \in T} \sum_{l \in L} \alpha_{lv} x_{lt} \geq 1, \quad \forall v \in V
\]  \hspace{1cm} (12)

\[
\sum_{t \in T} \sum_{l \in L} \beta_{le} x_{lt} \geq 1, \quad \forall e \in E
\]  \hspace{1cm} (13)

Constraints (12) assure that each station is covered by at least one train service. Constraints (13) indicate that each section is covered by at least one train service.

4.6.2 Passenger load in sections
For each section of the line, the transport capacity should not be less than the passenger load
of the section in both directions. That is, the passenger load of each section in each direction
should be satisfied by the combined transport capacity supplied by all selected train services.
In addition, in order to ensure the comfortableness of passengers and deal with the
occasional large passenger flow on the line, the operator of an urban rail transit line usually
holds a transport capacity surplus \( \delta \) when designing the train service scheme. We have:

\[
(1 - \delta) \sum_{t \in T} \sum_{l \in L} \beta_{le} r_{lt} y_{ltf} \geq \max_{e \in E} \{ p_{ed} \}, \quad \forall v \in E
\]  \hspace{1cm} (14)

4.6.3 Operational rules on train services
In the daily operation of an urban rail transit line, too many train services operated on the
line, a service with different train sizes and a service with a very low frequency will not
only increase the operation complexity of the line, but also not be conducive to the travel
experience of passengers. Hence, the number of operated train services, the number of train
sizes on a service and the minimum frequency of each service are restricted as follows:

\[
\sum_{t \in T} \sum_{l \in L} x_{lt} \leq \Omega
\]  \hspace{1cm} (15)

\[
\sum_{t \in T} x_{lt} \leq 1, \quad \forall l \in L
\]  \hspace{1cm} (16)

\[
\sum_{t \in T} \sum_{f \in F} y_{ltf} \geq \varphi \sum_{t \in T} x_{lt}, \quad \forall l \in L
\]  \hspace{1cm} (17)

\[
\sum_{f \in F} y_{ltf} \leq |F| x_{lt}, \quad \forall l \in L, \forall t \in T
\]  \hspace{1cm} (18)

Constraints (15) ensure that the number of train services operated on the line does not
exceed the upper limit \( \Omega \). Constraints (16) require that each service can choose at most one
type of train size. Constraints (17) assure that the frequency of a train service is not less
than the minimum value \( \varphi \) if this service is operated. Constraints (18) are the relationship
between variables \( x_{lt} \) and \( y_{ltf} \), indicating that the trains with size \( t \) can be operated on
service \( l \) only if this train size is adopted on the service.

4.6.4 Capacity of stations and sections

Generally, only a few stations on an urban rail transit line have turn-back facilities, and there may also be direction limit on the switch operation at each turn-back station. In addition, due to the track layout and other infrastructure facilities at turn-back stations, the number of trains that can be switched at each turn-back station in each direction within the study period is restricted by an upper limit \( s_{\text{std}} \). Thus, the capacity constraint of turn-back stations in the line can be formulated as follows:

\[
\sum_{l \in L} \sum_{t \in T} \sum_{f \in F} Y_{ltf} \leq s_{\text{std}} \quad \forall v \in V, \forall d \in D
\]  

(19)

Influenced by the train control system, on an urban rail transit line, the tracking headway between two adjacent trains to the same direction cannot be lower than a specified minimum value. Thus, during the study period, the frequency of trains in each section of the line should not be greater than the maximum frequency of the section, i.e. \( b_{e_{\text{max}}} \). Besides, in order to avoid passengers from waiting too long at some stations, the maximum headway in each section are limited when designing the train service scheme for an urban rail transit line. That is, during the study period, the frequency in each section of the line should not be less than the minimum frequency of the section, i.e. \( b_{e_{\text{min}}} \). Thus, the minimum and maximum frequency in sections of the line are satisfied by:

\[
b_{e_{\text{min}}} \leq \sum_{l \in L} \sum_{t \in T} \sum_{f \in F} \beta_{e} y_{ltf} \leq b_{e_{\text{max}}}, \quad \forall e \in E
\]  

(20)

4.6.5 Valid inequalities

In this paper, we introduce 0-1 variable \( y_{ltf} \) to indicate whether the \( f \)th train with size \( t \) on service \( l \) is operated. Thus, the frequency of trains with size \( t \) on service \( l \) is equal to the sum of a group of \( y_{ltf} \), i.e. \( \sum_{f \in F} y_{ltf} \). This will lead to a large quantity of feasible combinations of \( y_{ltf} \) under the same frequency for any service \( l \) and any train size \( t \), heavily aggravating the symmetry in the model. For instance, if in total 10 trains of size \( t \) is operated on service \( l \) (i.e. \( \sum_{f \in F} y_{ltf} = 10 \)) and the maximum frequency of trains is 20 (i.e. \( |F| = 20 \)), the number of feasible combinations of \( y_{ltf} \) is \( C_{20}^{10} \).

The symmetry of variable \( y_{ltf} \) can be easily broken by requiring that the \( (f+1) \)th train with size \( t \) on service \( l \) can be operated only if the corresponding \( f \)th train with the same size is operated. These valid inequalities are formulated as follows:

\[
y_{ltf} \geq y_{lt,f+1}, \quad \forall f \in [1,|F| - 1], \forall l \in L, \forall t \in T
\]  

(21)

4.6.6 Domain of variables

The integrity requirement of decision variables \( x_{lt} \) and \( y_{ltf} \) are provided by:

\[
x_{lt} = 0 \text{ or } 1, \quad \forall l \in L, \forall t \in T
\]

(22)

\[
y_{ltf} = 0 \text{ or } 1, \quad \forall l \in L, \forall t \in T, \forall f \in F
\]  

(23)

Now, we can completely formulate the train service design problem (TSD) in urban rail transit lines as a mixed integer nonlinear programming model M1 to minimize the objective function (6) with Constraints (7) to (23).
5 Linearization and Simplification of Model

5.1 Model Linearization

In model M1, there is only one nonlinear item $\sum_{t \in T} \sum_{f \in F} y_{taf}$ in Constraints (8). Indeed, it is a continuous variable $w_{nj}^{ij}$ multiplied by a 0-1 variable $y_{taf}$ which can be easily linearized by many existing techniques. Here, we present a novel linearization technique for this nonlinear item by utilizing the characteristics of the model. Recall that in our model the frequency of each train service is discretized from 1 to $|F|$, and variable $y_{taf}$ denotes whether the $f$th train with size $t$ is operated on service $l$. That is, the frequency of the $f$th train with size $t$ on service $l$ is just 1. Due to that, each boarding arc $a$ originating from station node $n$ in the service network could be duplicated by $|T| \times |F|$ times. The frequency of the duplicated arc corresponding to the $f$th train with size $t$ on boarding arc $a$ is only 1. A new non-negative contiguous variable $k_{taf}^{ij}$ is defined to represent the number of passengers of OD pair $(i,j)$ on the arc with respect to the $f$th train with size $t$ on boarding arc $a$. Thus, Constraints (8) can be linearized by the following constraints.

\begin{equation}
\kappa_{taf}^{ij} \leq \frac{2}{n} w_{n}^{ij}, \quad \forall t \in T, \forall f \in F, \forall a \in A_{n+} \cap A_{1\theta_{ij}}, \forall n \in N_{1}, \forall i \in V, j \in V
\end{equation}

\begin{equation}
\kappa_{taf}^{ij} \leq M_{1} y_{taf}, \quad \forall t \in T, \forall f \in F, \forall a \in A_{n+} \cap A_{1\theta_{ij}}, \forall n \in N_{1}, \forall i \in V, j \in V
\end{equation}

\begin{equation}
k_{taf}^{ij} = \sum_{t \in T} \sum_{f \in F} k_{taf}^{ij}, \quad \forall a \in A_{n+} \cap A_{1\theta_{ij}}, \forall n \in N_{1}, \forall i \in V, j \in V
\end{equation}

\begin{equation}
k_{taf}^{ij} \geq 0, \quad \forall t \in T, \forall f \in F, \forall a \in A_{n+} \cap A_{1\theta_{ij}}, \forall n \in N_{1}, \forall i \in V, j \in V
\end{equation}

Constraints (24) are the disaggregation representation of the optimal strategy requirement. These constraints take the same effects as Constraints (8). However, they are purely linear as the corresponding frequency is 1. Constraints (25) specify the relationship between variables $k_{taf}^{ij}$ and $y_{taf}$. There, $M_{1}$ is a large positive constant. For each OD pair $(i,j)$, it can take the volume $q_{ij}$ of the pair. Constraints (26) compute the number of passengers of OD pair $(i,j)$ on boarding arc $a$ by summing that on all the associated duplicated arcs. Constraints (27) are the non-negative requirements of variables $k_{taf}^{ij}$.

Through the above model linearization, the nonlinear train service design model M1 can be transformed into a mixed integer linear programming model M2 to minimize the objective function (6) and satisfy Constraints (7) and (9) to (27).

5.2 Model Simplification

5.2.1 Service network simplification

In our computational experiments, the model size and computation time are strongly influenced by the size of the service network. Observe that train services can be operated only between turn-back stations. Meanwhile, in each direction, the average waiting time of passengers at an intermediate station between two adjacent turn-back stations denoted as $tv_{1}$ and $tv_{2}$ ($tv_{2}$ is in front of $tv_{1}$ in the corresponding direction) is the same as that at turn-back station $tv_{1}$. Thus, we develop a method to reduce the size of the service network without losing of the solution accuracy of the model. The service network simplification
method works as follows:

Step1: In the service network, the station nodes and service nodes with respect to the stations without turn-back facilities are removed. Only the station nodes and service nodes related to turn-back stations are remained. Following the rules, the service network in Figure 3 is simplified as a smaller network shown in Figure 4.

Step2: The origin, destination and volume of OD pairs in the original service network are adjusted and aggregated in the simplified service network as follows:

(i) If the origin and destination of an OD pair are both turn-back stations, the origin and destination station node of this pair in the simplified network remain unchanged.

(ii) If the origin of an OD pair is not a turn-back station, the origin station node of this pair is set as the nearest turn-back station node behind the travel direction of this pair in the simplified network.

(iii) If the destination of an OD pair is not a turn-back station, the destination station node of this pair is set as the nearest turn-back station node in front of the travel direction of this pair in the simplified network.

(iv) After the adjustment of origin and destination, the volume of aggregated OD pairs in the simplified network is determined according to the volume of OD pairs in the original network. For instance, in the upward direction of Figure 4, the relationships between the volume of the aggregated OD pairs \( \{ q_{ij}' \mid i \in N_1, j \in N_1' \} \) and that of the original OD pairs \( \{ q_{ij} \mid i \in V, j \in V \} \) are as follows:

\[
q_{14}' = q_{12} + q_{13} + q_{14} + q_{24} + q_{34}
q_{47}' = q_{35} + q_{46} + q_{47} + q_{56} + q_{57} + q_{67}
q_{17}' = q_{15} + q_{16} + q_{17} + q_{25} + q_{26} + q_{27} + q_{35} + q_{36} + q_{37}
\]

To distinguish the simplified service network from the original one, additional notation shown in Table 2 is introduced.
After the service network simplification, the linear train service design model M2 can be reduced as a smaller size linear model M3 as follows:

\[
\begin{align*}
\text{min} & \sum_{i \in N} \sum_{f \in F} c_{ij} x_{ij} + \sum_{i \in N} \sum_{f \in F} c_{ij} y_{ij} + \frac{a}{n} \sum_{i \in N} \sum_{j \in N} w_{ij}
\end{align*}
\]

s.t. Constraints (11)-(22)

\[
\sum_{a \in A_{ns}} k_a - \sum_{a \in A_{ns}} k_a = \begin{cases} q_{ij}, & \lambda_n = i, \ n \in N'_i, \forall i \in N'_i, j \in N'_j \\
-q_{ij}, & \lambda_n = j, \ n \in N'_j, \forall i \in N'_i, j \in N'_j \\
0, & \text{else}
\end{cases}
\]

\[
k_{ij} \leq \frac{2}{M} w_{ij}, \ \forall t \in T, \forall f \in F, \forall a \in A_{ns}, \forall n \in N_i, \forall i \in N_i, j \in N_j
\]

\[
k_{ij} \leq M_2 y_{ij}, \ \forall t \in T, \forall f \in F, \forall a \in A_{ns}, \forall n \in N_i, \forall i \in N_i, j \in N_j
\]

In Constraints (31), the big-M parameter \( M_2 \) can be valued as \( q_{ij} \) for each OD pair.

### 5.2.2 OD pair aggregation

In model M3, variables \( w_{ij}, k_{ij}, k_{atf} \) and constraints with respect to passenger assignment are generated for each OD pair in the service network. The number of OD pairs and passenger assignment constraints increases at a square speed with the number of turn-back stations in the service network, leading to a rapid increase in the scale of the model. Consider that only the waiting time of passengers is influenced by train services. Meanwhile, under the optimal strategy-based passenger assignment, the trip of single OD pair does not impact the computation of the total waiting time of passengers. To further simplify the model, we refer to the OD pair aggregation method of Spiess (1989) to process the OD pairs in the service network. There, variables \( w_{ij}, k_{ij}, k_{atf} \) and passenger assignment constraints only need to be generated for each group of OD pairs to each destination station node in the service network. The OD pair aggregation method is implemented as follows:

**Step1:** Aggregate all OD pairs in the service network into groups of OD pairs such that
the OD pairs in each group go to the same destination station node. For destination station node \( i \) and station node \( n \) in the service network, let \( \mu_n^i \) either be the total number of passengers from all other station nodes to \( i \) if \( n = i \), or the number of passengers from station node \( n \) to \( i \) if \( n \neq i \). Thus, the relationship between \( \mu_n^i \) and \( q_{ij}^i \) is as follows:

\[
\begin{cases}
- \sum_{j \neq i} q_{ij}^i, & n = i, \forall i \in N'_t, \forall n \in N'_t \\
q_{ij}^i, & n \neq i, \forall i \in N'_t, \forall n \in N'_t
\end{cases}
\]

Step2: Replace variables \( w_{ij}^l \), \( k_a^l \) and \( \kappa_{a,f}^l \) with \( w_{ij}^l \), \( k_a^l \) and \( \kappa_{a,f}^l \), respectively. Here, \( w_{ij}^l \) indicates the waiting time of all the passengers to destination station node \( i \) at station node \( n \). \( k_a^l \) represents the number of all the passengers to destination station node \( i \) on boarding arc \( a \). \( \kappa_{a,f}^l \) denotes the number of all the passengers to destination station node \( i \) on the arc with respect to the \( f \)th train with size \( t \) on boarding arc \( a \).

Through the OD pair aggregation, model M3 can be further simplified as follows:

\[
\min \sum_{t \in T} \sum_{e \in E_t} \sum_{f \in F} \frac{c_{ej}^t y_{jtf}^T}{h} + \sum_{t \in T} \sum_{e \in E_t} \sum_{f \in F} c_{ej}^t y_{jtf}^T + \frac{\sigma}{60} \sum_{t \in T} \sum_{e \in E_t} \sum_{n \in N_1} w_{ij}^l
\]

s.t. Constraints (11)-(22)

\[
\sum_{a \in A_{nt}} k_a^l - \sum_{a \in A_{nt}} k_a^l = \begin{cases}
\mu^i_n, & \forall n \in N'_t, \forall i \in N'_t \\
0, & \forall n \in N'_t, \forall i \in N'_t
\end{cases}
\]

(38)

\[
k_{a,f}^l \leq \frac{2}{h} w_{ij}^l, \forall t \in T, \forall f \in F, \forall a \in A_{nt}, \forall n \in N'_t, \forall i \in N'_t
\]

(39)

\[
k_{a,f}^l \leq M_3 y_{t,f}^l, \forall t \in T, \forall f \in F, \forall a \in A_{nt}, \forall n \in N'_t, \forall i \in N'_t
\]

(40)

\[
k_a^l = \sum_{t \in T} \sum_{f \in F} k_{a,f}^l, \forall a \in A_{nt}, \forall n \in N'_t, \forall i \in N'_t
\]

(41)

\[
k_a^l = 0, \forall a \in A_{nt}, \forall n \in N'_t, \forall i \in N'_t
\]

(42)

\[
k_{a,f}^l \geq 0, \forall t \in T, \forall f \in F, \forall a \in A_{nt}, \forall n \in N'_t, \forall i \in N'_t
\]

(43)

\[
w_{ij}^l \geq 0, \forall n \in N'_t, \forall i \in N'_t
\]

(45)

The big-M parameter \( M_3 \) of Constraints (40) can take the value of \( \sum_{f \in N'_t, f \neq i} q_{ij}^i \) for each destination station node \( i \).

It is worth noting that the two model simplification techniques including the service network simplification and OD pair aggregation are independent. They can be used either separately or unitedly to simplify model M2. The order of the two simplification processes are not fixed. For comparison, we call model M2 simplified only by the OD pair aggregation as model M4, and model M2 simplified by both techniques as model M5.

6 Computational Experiments

In this section we describe our computational experiments on the (planned) Chongqing Rapid Rail Transit Line 26 in China. The proposed approaches are coded by MATLAB R2016a and CPLEX 12.8 is invoked to solve the optimization models. We run all experiments on a PC with Inter Core i7-7700 3.6 GHz CPU and 16 GB RAM.
6.1 Test Line and Parameter Setting

The total length of the test line with 20 stations and 19 sections is 121.7 km as shown in Figure 5. Stations are numbered from $v_1$ to $v_{20}$ along the upward direction. There are four turn-back stations namely station $v_1$, $v_6$, $v_{17}$ and $v_{20}$ on the line. Station $v_1$ can only switch the running direction of trains from downward to upward. Station $v_{20}$ can only switch the running direction of trains from upward to downward. Different from station $v_1$ and $v_{20}$, station $v_6$ and $v_{17}$ can reverse the running direction for trains from both directions. The capacity of turn-back stations in each direction are listed in Table 3. It should be mentioned that in addition to the turn-back track, the depot entrance and exit track can also reverse the running direction for trains during the operation period.

Six candidate train services on the line can be generated according to the layout of turn-back stations. The turn-back stations and round-trip time of each candidate train service are provided in Table 4. In the Table, the number in each cell represents the round-trip time of the train service formed by the turn-back stations in the row and column of the cell.

There are three types of train sizes that can be operated on the line, including 4-car trains, 6-car trains and 8-car trains. The relevant parameters of train sizes are described in Table 5.

Other parameters of the test line are in Table 6.
### 6.2 Instances Generation

In order to analyse the performance of the proposed model and the effectiveness of the two model simplification methods, 15 realistic instances based on the test line are constructed to test the performance of model M2, M3, M4 and M5.

We use three different scales of passenger demand at the morning rush hour from 8:00 to 9:00 in the initial, immediate and long-term planning horizon of the test line, as displayed in Figure 6. Among them, the total number of passengers of Figure 6(a), 6(b) and 6(c) is 87973, 98240 and 111136, respectively. For convenience, let $OD = 1,2,3$ represent the three scenarios of passenger demand in Figure 6. Besides, for each scenario of passenger demand, the maximum allowable number of train services $\Omega$ is set from 1 to 5 (i.e. $\Omega = \{1, \ldots, 5\}$). Thus, in total 15 instances are obtained to test the proposed approaches.

![Figure 6: Three scenarios of passenger demand](image_url)

### 6.3 Results

#### 6.3.1 Effectiveness of model simplification methods

We first compare the scale of the four configured models listed in Table 7. There, column 2 and 3 are the number of nodes and arcs in the associated service network, respectively. The last three columns are the number of 0-1 variables, continuous variables and constraints in the model, respectively. As shown, the scale of service network and scale of model in M2 are the largest. It has millions of variables and constraints. The size of service network in model M3 which is simplified by the layout of turn-back stations is obviously reduced. The number of continuous variables and constraints decrease significantly too. Model M4
has the same service network size as model M2. However, the OD pair aggregation method provides a smaller size to model M4 compared with model M2. The size of model M4 is between that of model M2 and M3. Model M5 simplified by the two methods has the same size of service network as model M3. But there are only near 5000 continuous variables and 8000 constraints in model M5. Thus, in terms of model scale, we have M2 > M4 > M3 > M5. In addition, all models have the same number of 0-1 variables. Because both of the two simplification methods only reduce the number of continuous variables $w_{ij}^{ij}$, $k_a^{ij}$ and $\kappa_{atf}^{ij}$ and the constraints which contain these continuous variables. The 0-1 variables $x_{it}$ and $y_{ltf}$ which determine the train service scheme are not simplified.

Based on the above comparison, we can conclude that the two model simplification methods including the service network simplification and OD pair aggregation can both

<table>
<thead>
<tr>
<th>Model</th>
<th>Nodes</th>
<th>Arcs</th>
<th>0-1 variables</th>
<th>Continuous variables</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2</td>
<td>168</td>
<td>408</td>
<td>378</td>
<td>3435200</td>
<td>4941789</td>
</tr>
<tr>
<td>M3</td>
<td>36</td>
<td>60</td>
<td>378</td>
<td>20224</td>
<td>23837</td>
</tr>
<tr>
<td>M4</td>
<td>168</td>
<td>408</td>
<td>378</td>
<td>171760</td>
<td>259813</td>
</tr>
<tr>
<td>M5</td>
<td>36</td>
<td>60</td>
<td>378</td>
<td>5056</td>
<td>8057</td>
</tr>
</tbody>
</table>

We then analyse the computational effectiveness of the configured models. The maximum running time of each model is limited to 4 h. The computational results are summarized in Table 8. In the Table, Columns 3-6 are the objective function value of the models. Columns 7-10 are the optimality gap between the best lower bound and upper bound. The last four columns are the computation time. As observed, model M2 can only find the optimal solution for three instances ($OD = \{1,2,3\}, \Omega = \{1\}$). For other instances, only feasible solutions with large gaps (39% to 56%) are obtained in 4 h. Contrarily, Model M3, M4 and M5 can obtain the optimal solution for all instances within the limited time. The computation time of the four models is consistent with the scale of the models. The solving speed of model M3 and M5 is quite fast with an average computation time less than 10 s. Due to the model scale, Model M4 is relatively difficult to solve and its average computation time is near 20 min.

<table>
<thead>
<tr>
<th>Instance</th>
<th>Objective/¥</th>
<th>Gap/%</th>
<th>Time/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>OD</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
</tr>
<tr>
<td>1</td>
<td>644286</td>
<td>644286</td>
<td>644286</td>
</tr>
<tr>
<td>2</td>
<td>881654</td>
<td>491088</td>
<td>491088</td>
</tr>
<tr>
<td>3</td>
<td>881654</td>
<td>476368</td>
<td>476368</td>
</tr>
<tr>
<td>4</td>
<td>881654</td>
<td>476368</td>
<td>476368</td>
</tr>
<tr>
<td>5</td>
<td>881654</td>
<td>476368</td>
<td>476368</td>
</tr>
<tr>
<td>Ave</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>718038</td>
<td>718038</td>
<td>718038</td>
</tr>
<tr>
<td>2</td>
<td>890171</td>
<td>536576</td>
<td>536576</td>
</tr>
<tr>
<td>3</td>
<td>890171</td>
<td>526337</td>
<td>526337</td>
</tr>
<tr>
<td>4</td>
<td>890171</td>
<td>526337</td>
<td>526337</td>
</tr>
<tr>
<td>5</td>
<td>890171</td>
<td>526337</td>
<td>526337</td>
</tr>
<tr>
<td>Ave</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on the above comparison, we can conclude that the two model simplification methods including the service network simplification and OD pair aggregation can both
effectively reduce the scale of the original model and improve the solution quality. When
the original model is simplified by one of the two methods alone, the service network
simplification has better effects in reducing the model scale and improving the solving
speed. Note that the OD pair aggregation also has a notable simplification effect. When the
two model simplification methods work together, the simplified model M5 has the smallest
model scale and the shortest computation time, which enables us to solve practical-sized
train service design problems of urban rail transit lines in extremely short time.

6.3.2 Comparison with single train service scheme
To testify the quality of the train service scheme we proposed, we compare the multiple
train services scheme obtained by model M5 with the single train service scheme frequently
designed by experience in practice. For simplicity, we only design train services in the long-
term planning horizon of the test line. The corresponding passenger demand is shown in
Figure 6(c). The two schemes are obtained as follows:

(i) Single train service scheme (STSS). In practice, the train service scheme of an urban
rail transit line is usually determined by using a full-length train service and a single
size according to the maximum passenger load of sections in the line. Under this rule,
the best single train service scheme of the test line can be obtained by setting $OD = 3$, $\Omega = 1$ and $T = \{8\}$ in model M5. The resulting train size and frequency of
the single full-length service are the 8-car train and 14 pairs of trains, respectively.

(ii) Multiple train services scheme (MTSS). To obtain the optimal train service scheme
of the test line, we can set $OD = 3$, $\Omega = 5$ and $T = \{4,6,8\}$ in model M5. The
obtained multiple train services scheme is depicted in Figure 7. As seen, three are 4
train services where 4-car trains and 8-car trains are used.

We first analyse the transfer of passengers under different train service schemes. For
scheme STSS, as a full-length train service is operated, passengers do not need to transfer
when traveling on the line. Regarding to scheme MTSS, some passengers need to transfer
at most twice when they travel. For example, the travel process of partial passengers from
station $v_{20}$ to station $v_1$ is $v_{20}$-Service 4-$v_{17}$-Service 3-$v_6$-Service 1 or Service 2-$v_1$,
and hence the number of transfers is two. However, only the passengers from stations $v_{18}$,
$v_{19}$ and $v_{20}$ to stations $v_1$, $v_2$, $v_3$, $v_4$ and $v_5$ need to transfer twice during their
journey. The number of these passengers is only 457, accounting for 0.41% of the total
number of passengers. We will indicate that the operator’s cost and passengers’ cost can be
reduced significantly while a tiny proportion of passengers has an inconvenient journey.

![Figure 7: Proposed multiple train services scheme](image)

Then we compare the frequency of trains in sections for different schemes are shown in
Figure 8. As indicated, scheme STSS has a single full-length service covering all stations
and sections of the line, i.e. 14 pairs of 8-car trains are operated on the whole line. In scheme
MTSS, 12 pairs of 4-car trains are operated in sections $e_1$ to $e_5$, 12 pairs of 4-car trains
and 8 pairs of 8-car trains serve sections $e_6$ to $e_{16}$, and 15 pairs of 4-car trains run in
sections $e_{17}$ to $e_{19}$. The frequency of trains in scheme MTSS is higher than that of scheme STSS in sections $e_6$ to $e_{19}$, where exist most of the passenger flows in the line, thus leading to a shorter waiting time for most of the passengers.

![Figure 8: Frequency of trains in sections](image)

We further analyse the match between capacity and demand in sections for both schemes. The capacity, demand and load factor in sections under the two schemes are displayed in Figure 9. As known, in scheme STSS, only 8-car trains are used and the frequency of trains is the same in each section of the line. Therefore, the capacity in each section is equal. However, the passenger load is obviously unbalanced in sections. The passenger load in the middle sections of the line is large while that in the two ends is small. This leads to the waste of capacity in the two ends of the line, which is not economical for the operator. On the contrary, for scheme MTSS, 4 services and 2 types of train sizes are used on the lines, such that different sections of the line can be more flexibly equipped with capacity. The capacity in sections $e_1$ to $e_5$ and $e_{17}$ to $e_{19}$ are smaller but match the passenger load better than scheme STSS, which can help to reduce the cost of operating the line. Note that lower capacity in sections does not necessarily mean lower frequency of trains in sections. Because in scheme MTSS, many small size trains are operated to increase the frequency of trains in sections so as to reduce the waiting time of passengers. As shown in Figure 8, the frequency of trains under MTSS in sections $e_1$ to $e_5$ is lower than that of STSS. But in sections $e_{17}$ to $e_{19}$, the frequency of trains in MTSS is higher than that of STSS.

Finally, we compare the objective function value under different schemes as summarized in Table 9. It can be seen from Table 9 that compared with scheme STSS, scheme MTSS reduces the total cost by 26.58%. Meanwhile, the total fixed cost of operators, the total operation cost of operators and the total waiting time cost of passengers are all decreased. Besides, through a close look at the composition of the total cost in scheme MTSS, we can find that the total operation cost $Z_2$ accounts for most of the total cost with a rate of 81.06%, while the total fixed cost $Z_1$ with a rate of 2.37% is the smallest part of the total cost. This is different from our empirical understanding that the fixed cost may account for the majority of the total cost in urban rail transit systems. Nevertheless, this difference reflects that we should save the operation cost as much as possible so as to control the total cost of operating multiple train services in urban rail transit lines.
In summary, in light of the above comparisons, we conclude that our approach which can flexibly design a train service scheme with multiple services and multiple train sizes is better than the empirical method frequently used in practice to design a single train service scheme. With our approach, although some passengers may experience a limited number of transfers, the capacity in sections can match the passenger load better. Meanwhile, the frequency of trains in sections can be increased as much as possible, reducing significantly of the operators’ cost and passengers’ waiting time cost. Thus, the proposed approach can be used to design practically acceptable train services in urban rail transit systems.

Figure 9: Capacity, demand and load factor in sections

Table 9: Objective function value of schemes

<table>
<thead>
<tr>
<th>Objective</th>
<th>STSS</th>
<th>MTSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total cost/ ¥</td>
<td>810600.2</td>
<td>595173.9</td>
</tr>
<tr>
<td>Total fixed cost of operators $Z_1$/ ¥</td>
<td>18145.8</td>
<td>14087.3</td>
</tr>
<tr>
<td>Total operation cost of operators $Z_2$/ ¥</td>
<td>681318.4</td>
<td>482457.2</td>
</tr>
<tr>
<td>Total waiting time cost of passengers $Z_3$/ ¥</td>
<td>111136.0</td>
<td>98629.4</td>
</tr>
</tbody>
</table>
7 Conclusions

This study proposes an optimization approach for the train service design in an urban rail transit line considering short-turn services and multiple train sizes. A service network is constructed based on a pool of candidate services generated in advance. By considering a series of operational and capacity constraints, the problem is formulated as a mixed integer nonlinear programming model to minimize the operators’ cost and passengers’ waiting time cost. After a model linearization, two simplification methods namely service network simplification and OD pair aggregation are used to simplify the linear model. Finally, the Chongqing Rapid Rail Transit Line 26 in China is used to test our approach.

Compared with the existing studies, the model we proposed is more flexible in the sense that it can be applied to the train service design of an urban rail transit line with multiple train services including either full-length or short-turn ones and multiple train sizes. Next, seldom former studies consider the travel process of passengers in case of multiple train services in urban rail transit systems. In this work, we propose a service network construction method based on a pool of candidate services generated in advance, and use the optimal strategy to assign passenger flows in the service network so as to well describe the travel process of passengers, such that the impact of multiple train services on the trip choices of passengers can be accurately analysed. Furthermore, a service network simplification and an OD pair aggregation are developed to simplify our model efficiently. Computational experiments show that the model with the two simplification techniques can be quickly solved to optimality by commercial solvers for typical urban rail transit lines. The obtained multiple train services scheme effectively reduces the operator’s cost and passengers’ waiting time cost compared to the single train service scheme frequently designed by experience in practice.

Acknowledgments

We are partially supported by the National Natural Science Foundation of China (No. 61603318, U1834209), and the National Key Research and Development Program of China (No. 2017YFB1200701).

References


