Transportation Research Part E 174 (2023) 103142

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

Synchronization of train timetables in an urban rail network: A bi-objective optimization approach

Jiateng Yin [a,b,*], Miao Wang [a], Andrea D'Ariano [c], Jinlei Zhang [b,*], Lixing Yang [a,b]

[a] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China
[b] School of Systems Science, Beijing Jiaotong University, Beijing, 100044, China
[c] Department of Civil, Computer Science and Aeronautical Technologies Engineering, Roma Tre University, Rome, Italy

ARTICLE INFO

Keywords:
Urban rail transit
Timetable optimization
Transfer synchronization
Passenger waiting time

ABSTRACT

As urban rail networks in big cities tend to expand, the synchronization of trains has become a key issue for improving the service quality of passengers because most urban rail transit systems in the world involve more than one connected line, and passengers must transfer between these lines. In contrast to most existing studies that focus on a single line, in this study, we focus on synchronized train timetable optimization in an urban rail transit network, considering the dynamic passenger demand with transfers as well as train loading capacity constraints. First, we propose a mixed-integer programming (MIP) formulation for the synchronization of training timetables, in which we consider the optimization of two objectives. The first objective is to minimize the total waiting time of passengers, involving arriving and transfer passengers. Our second objective is a synchronization quality indicator (SQI) with piecewise linear formulation, which we propose to evaluate the transfer convenience of passengers. Subsequently, we propose several linearization techniques to handle the nonlinear constraints in the MIP formulation, and we prove the tightness of our reformulations. To solve large-scale instances more efficiently, we also develop a hybrid adaptive large neighbor search algorithm that is compared with two benchmarks: the commercial solver CPLEX and a metaheuristic. Finally, we focus on a series of real-world instances based on historical data from the Beijing metro network. The results show that our algorithm outperforms both benchmarks, and the synchronized timetable generated by our approach reduces the average waiting time of passengers by 1.5% and improves the connection quality of the Beijing metro by 14.8%.

1. Introduction

Urban rail transit is playing an increasingly significant role worldwide because of its characteristics of high comfort, carbon efficiency, and convenience for passengers. Owing to these advantages, the number of passengers traveling with urban rail transit has reached 10 million per day in metropolises such as Beijing, Shanghai, and Tokyo (Wang et al., 2018). The management of such complex systems, involving route planning, timetable optimization, vehicle circulation, and crew scheduling, is an extremely difficult task for rail operators (Ibarra-Rojas and Rios-Solis, 2012; Yang et al., 2016). In urban rail transit systems, all train services are operated according to a predetermined timetable that is publicly available for passengers. Thus, the optimization of train timetables that provide high-quality services to passengers is particularly important for high-density urban metro networks.

* Corresponding author.
E-mail addresses: jtyin@bjtu.edu.cn (J. Yin), zhangjinlei@bjtu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.tre.2023.103142
Received 30 August 2022; Received in revised form 22 February 2023; Accepted 27 April 2023
Available online 8 May 2023

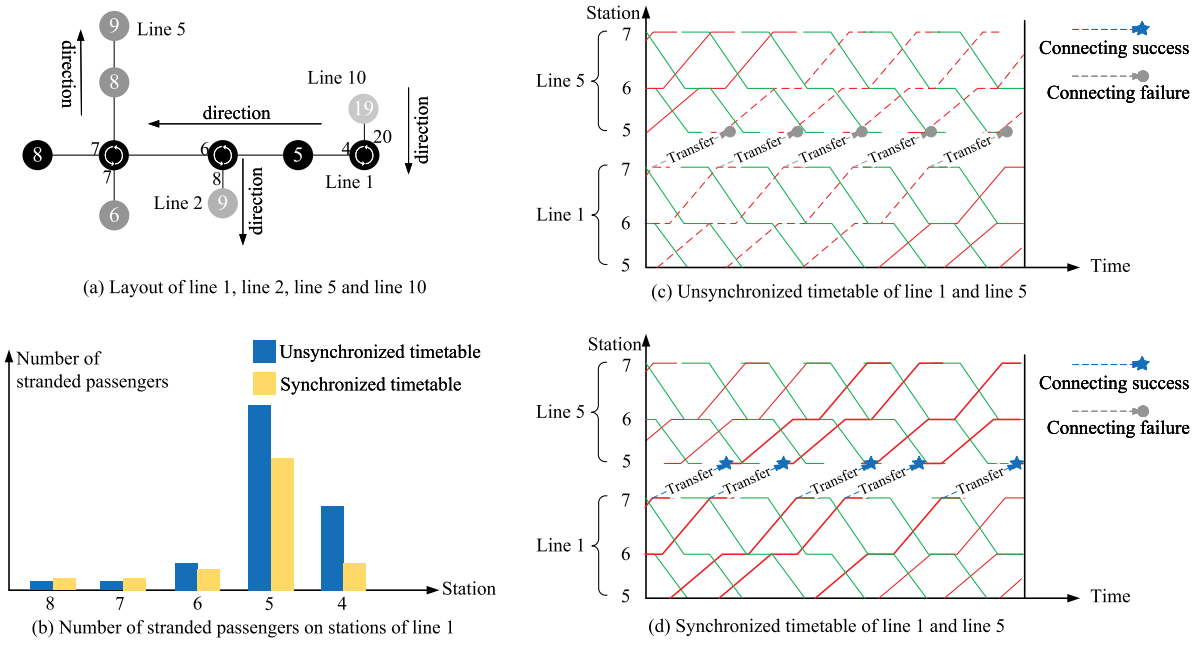


Fig. 1. Illustration of Beijing metro network, with and without timetable synchronization.

The expansion of the urban rail networks enhances passenger accessibility. However, the passengers may have to transfer more than once among different lines because there may be no direct route between their origins and destinations. According to the statistics of the Beijing metro, passengers transfer more than twice on average during a single trip (CAMET, 2021). Consider the example in Fig. 1(a). The transfer passengers may have to wait for a longer time as they will miss the connected trains (as shown in Fig. 1(c)). By considering the specific transfer time between the connecting lines, a well-designed timetable (as shown in Fig. 1(b) and (d)) can provide "smooth" transfers to passengers and thus considerably improve their traveling convenience. Therefore, the synchronization of timetables in urban rail networks has received increasing attention from urban rail managers (Liu et al., 2021). Nevertheless, there are few studies that focus on the passenger demand-oriented timetable synchronization, considering the complexity of transfer passenger flows and limited train capacities. The difficulty of the problem is increased due to the following two main reasons. First, the modeling of passenger flows is very complicated owing to the dynamic characteristics of passenger demand and limited train capacities (Niu et al., 2015). When considering an entire network comprising several connected lines, this problem becomes even more challenging because the volume of transfer passengers is also affected by the timetables of the feeder lines. Second, the synchronization of train timetables may also increase the waiting time of passengers without transfers. Therefore, qualitatively modeling the transfer convenience in a large transit network with high-frequency services and analyzing the relationship between transfer convenience and waiting time of passengers are still challenging issues (Wang et al., 2021).

To address the problems mentioned above, we propose a synchronization quality index (SQI) to qualitatively measure the transfer convenience of a synchronized train timetable. We explicitly model the flow of passengers in the network considering the transfer activities and limited train capacity, which enables quantification of the exact waiting time of the passengers involved. Then, we construct a mixed-integer programming (MIP) model with two objectives, that is, SQI and the waiting time of traveling passengers. Because of the non-convexity of the SQI function, we propose a series of linearization techniques and prove that our reformulated SQI function is locally ideal, that is, it provides the strongest LP bound. Furthermore, we develop a hybrid solution algorithm based on an adaptive large neighborhood search (ALNS) to solve large-scale instances more efficiently.

1.1. Literature review

A wide range of studies have been devoted to train timetabling problems in urban rail systems, with the objectives of improving the service quality of passengers and reducing the operational costs (Wong et al., 2008; Shafahi and Khani, 2010; Shi et al., 2018; Yin et al., 2019). We now review relevant studies in two categories: train timetabling for a single line and for multiple lines.

1.1.1. Timetable optimization for a single metro line

Currently, most studies on the train timetabling optimization are devoted to a single metro line. These studies consider the time-varying characteristics of passenger demand in urban rail lines, and aim to design demand-oriented irregular timetables with improved service quality. For example, considering the time-variant passenger demands, Niu and Zhou (2013) constructed a nonlinear integer programming model for the train timetabling of a metro line under oversaturated conditions, where some

passengers may have to wait for multiple trains owing to the limited loading capacity during peak hours. In their formulation, the concept of the *passenger loading time window* was introduced to model the waiting time of passengers analytically. Barrena et al. (2014) presented two mixed-integer nonlinear programming models to minimize the average passenger waiting time under dynamic passenger demand, and an ALNS algorithm was developed to solve the models. Sun et al. (2014) formulated three MIP models for the optimization of the train timetable, which were used for unlimited capacity, limited capacity, peak/off-peak, and limited capacity condition. To minimize the energy consumption and passenger waiting time simultaneously, Yin et al. (2017) developed two mixed-integer linear programming (MILP) models based on the space–time network formulations. A Lagrangian-based heuristic algorithm was developed to solve large-scale instances more efficiently.

More recently, a few studies have combined the train timetabling problem with other decisions, such as rolling stock circulation and passenger control strategies. For example, Wang et al. (2018) proposed a mixed-integer nonlinear programming model to simultaneously optimize the rolling stock circulation plan and train timetables for an urban rail line. Several model approximation techniques and iterative frameworks have been developed to solve the integrated model more efficiently. To avoid congestion at platforms for an oversaturated metro line with time-varying demands, Shi et al. (2018) proposed an integer programming (IP) model to collaboratively optimize the train timetable and passenger flow control strategies. A two-step iterative algorithm was developed to efficiently solve practical instances.

Overall, the above studies focused on the train timetabling problem for a single metro line under dynamic passenger demand, and the results were shown to be effective for reducing the passenger waiting time. Nevertheless, as urban rail systems commonly involve more than one line that are connected to each other, these studies on single lines cannot model the situation of transfer passengers and thus cannot be directly applied to a transit network (Yin et al., 2021).

1.1.2. Timetable optimization for urban transit networks

In recent years, an increasing number of studies have been devoted to the vehicle timetabling problem, considering the entire transit network (Chai et al., 2023). Some researchers have studied the synchronization of timetables to maximize the number of simultaneous arrivals of vehicles or minimize the waiting time of transfer passengers (Wu et al., 2015; Guo et al., 2017; Chu et al., 2019). For instance, considering the timetable synchronization of bus networks, Ceder et al. (2001) constructed an MILP model to simultaneously maximize the number of buses arriving at the station. Shafahi and Khani (2010) proposed two MIP models to minimize the transfer waiting time of passengers, where the first model determines the departure times at the first station, and the second model adds the extra stopping time of buses at transfer stations. In these studies, the authors mainly focused on transfer stations, while the demands of passengers were not explicitly considered.

For urban rail transit networks, Wong et al. (2008) first proposed an MIP model that aims to minimize the transfer waiting time of passengers, while the capacity of the trains was assumed to be sufficient at any time to accommodate all the waiting passengers. Subsequently, a series of mathematical formulations were developed for the optimization of different objectives. For example, in Li et al. (2019), a new satisfaction function of transfer passengers was defined based on three parameters, i.e., the passengers' expectation, tolerance, and dissatisfaction on "just miss" events. A real-world experiment was conducted on the Shanghai Urban Rail Transit Network. The results show that, compared with the minimization of total transfer waiting time, maximizing the proposed satisfaction function achieves a better performance in reducing "long wait" and "just miss" events. Cao et al. (2019) proposed an MIP model to maximize the number of synchronized meetings for urban transit networks. Instead of fixed values for passenger transfer time, they considered that the walking time of transferring passengers is flexible within a given range. The model was solved by a genetic algorithm embedded in a local search strategy.

A few studies have also addressed the timetable coordination problem of the first or last train in urban rail networks, where the objective is to minimize the number of passengers who are unable to board the first/last train (Kang and Zhu, 2016). Considering that the demand for upstream and downstream passenger flow considerably varies in the early morning, Guo et al. (2016) studied the synchronization of train timetables for the first train, where the importance of stations and lines is set as different costs according to the passenger demand. With the consideration of dynamic passenger demand during peak-hours, Yin et al. (2021) defined the concept of "congestion level" as the objective function and they proposed an MILP model to generate a coordinated train timetable that minimizes the congestion of transfer stations. A decomposition-based adaptive large neighborhood search algorithm was designed to solve the model.

In urban rail networks, the objectives associated with different passengers differ. For example, the passengers who need to transfer multiple times are concerned about the transfer convenience, whereas the other passengers may prefer shorter waiting times or lower congestion. Therefore, recent studies have started considering multi-objective optimization models for the train timetabling in urban rail networks. For instance, a bi-objective mix-integer quadratic programming (MIQP) model was proposed by Tian and Niu (2018) to maximize the number of connected trains and minimize the total transfer waiting time. Nevertheless, the number of transfer passengers between trains was assumed to be known in advance, and the trains were assumed to have sufficient capacity to accommodate all the transfer passengers. Liu et al. (2020) constructed a multi-objective mixed-integer nonlinear programming model that comprehensively considers the utilization of train and the passenger flow control strategies at platforms. A Lagrangian relaxation-based algorithm was proposed to obtain a trade-off solution among the maximum train services, minimum number of waiting passengers, and minimum amplitude of passenger flow control actions. However, the arriving passenger flows in their study were considered as static values for each station. Wang et al. (2020) constructed an MILP model with a bi-objective function to minimize the waiting time of both non-transfer passengers and transfer passengers, and the number of passengers who missed the transfers. A genetic algorithm was developed to solve the MILP model. In their study, the oversaturated situations were not allowed, as their objective was only to consider waiting passengers and transferring passengers for the first arriving train. In addition, the boarding/alighting processes of passenger flows were not explicitly considered in their formulation. .

Table 1
Summary of relevant studies for the timetabling of urban rail systems.

Publications	Passenger demand	Train capacity	Objective	Line/network	Model	Solution method
Ceder et al. (2001)	Static	Unconsidered	Number of simultaneous arrivals	Network	MILP	Heuristic algorithm
Guo et al. (2016)	Unconsidered	Unlimited	Transfer synchronization events	Network	MILP	Heuristic algorithm
Tian and Niu (2018)	Static	Unconsidered	Transfer waiting time and number of connections	Network	MIQP	Heuristic algorithm
Shafahi and Khani (2010)	Static	Unlimited	Transfer waiting time	Line	MIP	Local search + CPLEX
Kang and Zhu (2016)	Static	Unlimited	Arrival time differences and number of missed trains	Network	MIP	Local search
Cao et al. (2019)	Static	Unlimited	Number of synchronized services	Network	MILP	Local search and genetic algorithm
Chu et al. (2019)	Static	Unlimited	Total travel time	Network	MILP	Heuristic algorithm
Li et al. (2019)	Static	Unlimited	Transfer waiting time	Network	MILP	Heuristic algorithm
Wong et al. (2008)	Piecewise constant	Unlimited	Transfer waiting time	Network	MIP	Heuristic + CPLEX
Shi et al. (2018)	Dynamic	Limited	Waiting time of passengers	Line	IP	Improved local search & CPLEX
Wang et al. (2020)	Dynamic	Limited	Total waiting time and number of failed transfer passengers	Network	MIP	Genetic algorithm and gray wolf algorithm
Yin et al. (2021)	Dynamic	Limited	Crowdedness of stations	Network	MILP	Decomposition-based adaptive large neighborhood search
This paper	Dynamic	Limited	Waiting time of network and synchronization quality	Network	MIP	Adaptive large neighborhood search

1.2. Contribution of this paper

In our study, we focused on the synchronization of train timetables in an urban rail network to improve the service quality of both transfer passengers and non-transfer passengers. We explicitly modeled the time-variant passenger flows and transfer passengers in the network by considering the limited train capacity. Table 1 compares the mathematical formulations and solution methods proposed in this study compared with those in recent literature. The unique contributions of this paper are summarized as follows.

(1) We propose an MIP formulation for the synchronization of train timetables in an urban rail network. Our formulation considers the optimization of a bi-objective function under time-dependent passenger demand. The first objective is an SQI with piecewise linear formulation, which we propose to evaluate the transfer convenience of passengers. The second objective is to minimize the total waiting time of passengers, including the waiting time of newly arriving passengers, stranded passengers, and transfer passengers.

(2) Owing to the nonlinearity of the MIP formulation, we propose several reformulation techniques by introducing auxiliary variables and lifting the formulation to higher dimensions. Regarding the reformulation of SQI, we also prove its tightness from a piecewise linear non-convex function to a convex function. The reformulations enable us to directly solve small or medium instances with commercial solvers, such as CPLEX.

(3) To solve large-scale instances more efficiently, we develop a hybrid ALNS algorithm by combining it with simulated annealing. We compare our algorithm with two benchmarks: the commercial solver CPLEX and a traditional metaheuristic genetic algorithm. The experimental results show that, for small-scale instances, our algorithm can obtain near-optimal solutions close to CPLEX in a very short time. For large scale instances that CPLEX cannot solve, our algorithm can still find good solutions within a reasonable time. For all the instances, our algorithm evidently outperformed the metaheuristic algorithm.

(4) A series of real-world experiments based on the data from the Beijing Metro are implemented to verify the effectiveness of our approach. The results show that, our optimization based approach can remarkably improve the synchronization quality by 14.8% and reduce the passenger waiting time by 1.5%, compared with the existing timetable in the Beijing Metro. We also derive a few marginal insights that help rail managers create a better timetable for urban rail networks.

The remainder of this paper is organized as follows. Section 2 presents a detailed problem statement and establishes the MIP model. We then reformulate the MIP model into a mixed integer quadratic programming model with linear constraints in Section 3. Section 4 describes the explicit steps used to solve this model. Two sets of numerical experiments are presented in Section 5. This paper concludes with Section 6.

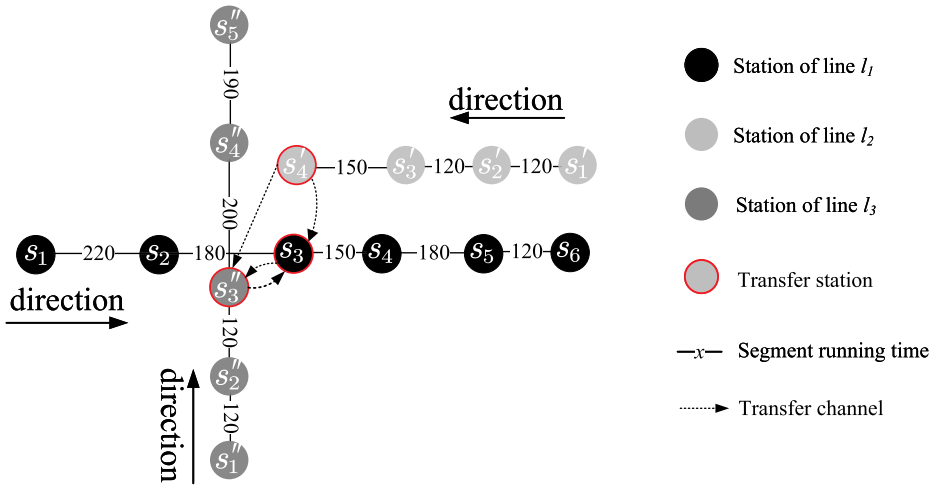


Fig. 2. An illustrative urban rail transit network with three connected lines.

2. Model formulation

In this section, we first describe the problem studied in this paper and then formulate the problem into an MIP model by introducing the model constraints and objective functions.

2.1. Problem statement

2.1.1. Network structure

We consider an urban rail transit network involving a set of lines, denoted by L. Each line $l \in L$ includes a set of stations $S_l = \{1, 2, \cdots, |S_l|\}$ and a set of trains $Q_l = \{1, 2, \cdots, |Q_l|\}$, which are to be scheduled between stations 1 and $|S_l|$. Let $(s', s) \in S^r$ denote the set of transfer stations, which indicates that passengers can transfer from station s' to station s. We also define $H_s = \{l \in L | s' \in S_{l'}, (s', s) \in S^r\}$ as the set of transfer feeder lines with respect to transfer station s, indicating that passengers from line l can transfer to station s. Note that the lines in real life are bidirectional tracks. In other words, each line is composed of both up-directional and down-directional lines. The above-defined sets involve the lines and stations in both directions. For convenience of description, hereafter, we use symbols q, q', and q'' to index the train services running on line l_1, l_2, and l_3, and symbols s, s', and s'' to denote the indexes of stations belonging to lines l_1, l_2, and l_3, respectively.

A toy example with three connected lines is presented in Fig. 2, where the network comprises line l_1 with six stations, line l_2 with four stations, and line l_3 with five stations. These three lines are connected by a transfer station. Therefore, according to the definition of the transfer feeder lines, we have: $H_{s_3} = \{l_2, l_3\}$, $H_{s'_4} = \varnothing$ and $H_{s''_3} = \{l_1, l_2\}$. Note that the transfer station is the last station of line l_2; therefore, its transfer feeder line set is actually an empty set.

2.1.2. Dynamic passenger demand

In urban rail transit networks, passenger demand is recognized to be time-dependent and station-dependent, and we can capture the arrival rate of passengers at different times and at different stations using historical data collected from the Auto Fare Collection (AFC) system (Liu et al., 2020). To model the time-dependent passenger demand in an urban rail network, we first discretize the time horizon into a set of time periods, that is, $T = \{T_1, T_2, \cdots, T_K\}$, where we use T_k to denote the kth time period, and T_1 and T_K to denote the beginning and end of the time horizon, respectively. Each time period represents a constant value of the time interval δ (e.g., 15 minutes). Then, we can denote a demand matrix $R = \{R_s(T_k) | T_k \in T, s \in S_l, l \in L\}$, where $R_s(T_k)$ represents the arrival rate, that is, the number of passengers arriving at station s per second during time period T_k. Here, the arrival rate $R_s(T_k)$ in each period is assumed to be constant in our study, similarly to other studies, for example, 15 min in Ying et al. (2020).

To model the number of passengers alighting at each station, we adopted a strategy similar to Huang et al. (2020). Specifically, we consider that the alighting rates of passengers at each station are given beforehand according to the historical AFC data. Thus, we use a passenger alighting rate vector $R^A = \{R^A_s | s \in S_l, l \in L\}$, where each element R^A_s represents the proportion of passengers who get off at station s to the number of passengers in-vehicle before the train arrives at station s. The alighting rate at the last station is 1, that is, no passenger stays in the vehicle after the train arrives at the last station. We denote another set $R^r = \{R^r_{s's} | (s', s) \in S^r\}$ to denote the set of transfer rates, where each element $R^r_{s's}$ represents the proportion of the number of transfer passengers to station s among the number of alighting passengers at station s'.

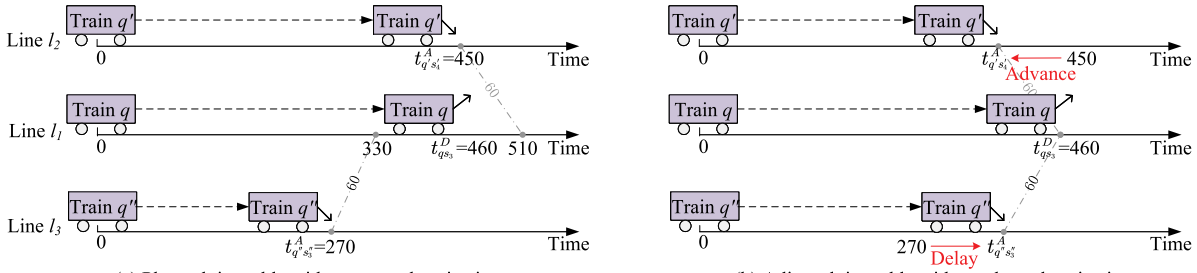


Fig. 3. An illustrative example of timetable synchronization.

2.1.3. Problem description and assumptions

In urban rail networks, the train timetabling problem aims to optimize the arrival/departure and dwell times of trains at stations, such that the service quality of passengers (QoS) is optimized. Nevertheless, the QoS value is different for the passengers with and without transfers. For each single line, the rail operator aims to create an irregular timetable, such that the passenger waiting time or traveling time is minimized (Hassannayebi et al., 2016). Meanwhile, when considering the whole network, a timetable that improves transfer convenience while providing good service for all the involved passengers is particularly favored by rail managers (Liu and Ceder, 2018). However, these two objectives, that is, the QoS for a single line and an entire network, may be contradictive in practical cases.

Consider the example in Fig. 3 (with the illustrative network is shown in Fig. 2). In the example, three trains q, q', and q'' are running on lines l_1, l_2, and l_3, respectively. All three trains depart from the first station at time $t = 0$. The dwell time of each train at each station is 30, the walking time of passengers in the transfer channel is 60 and the departure headway at the first station is 180. As shown in Fig. 3, train q' and train q'' arrive at the transfer station at time 450 and 270, respectively, while train q departs from the transfer station at time 460. Obviously, the transfer between q' and q is unsuccessful because the actual waiting time of the passengers is $\Delta t = 460 - 450 - 60 = -50$, which means that train q has already departed when the transfer passengers arrive at the platform. For the transfer between q'' and q, the passenger waiting time is $\Delta t = 460 - 270 - 60 = 130$. The quality of transfer connection can be improved by adjusting the departure times of q' and q'', which however may lead to an increase in the waiting time of passengers traveling on lines l_2 and l_3.

In summary, our study aims to construct a bi-objective optimization model to generate a synchronized train timetable that achieves a trade-off between passenger waiting time and transfer convenience in an urban rail network. Our study is based on the following assumptions.

Assumption 1. The transfer walking time of passengers is pre-given and uniform for the same transfer channel. Similar assumptions were adopted for Shafahi and Khani (2010) and Tian and Niu (2018).

Assumption 2. The train capacity provided within the planned time horizon can satisfy the total passenger demand, that is, all passengers can successfully board the train at the end of the planning time horizon.

Assumption 3. Passengers are assumed to be more inclined to choose trains that take less waiting time and would not leave the platform owing to congestion (Guo et al., 2017). All the passengers will get off the train at the last station.

2.2. Mathematical formulation

In this section, we develop an MIP model with a nonlinear objective function to solve the timetable synchronization problem. First, we describe the notation and decision variables used in the formulation. We then construct a series of operational constraints, involving the constraints with respect to trains and passengers. Finally, we give the definition of bi-objective function with respect to the transfer convenience and passenger waiting time.

2.2.1. Notations

The relevant notations are listed below to describe and model the problem more conveniently (see Tables 2 and 3).

2.2.2. Timetable related constraint

In practice, the departure time of the first train has a critical impact on the distribution of cumulative passengers. Therefore, rail operators typically require the departure time of the first train at the first station to be within an appropriate range (Kang and Zhu, 2016). Thus, we have the following constraints:

$$\underline{t}_{q,l} \leqslant t^D_{qs} \leqslant \bar{t}_{q,l}, \forall q = 1, s = 1, l \in L \tag{1}$$

Table 2
Parameters and notations.

Symbol	Description
$L = \{1, 2, \cdots, \|L\|\}$	Set of lines, where $\|L\|$ denotes the total number of lines in transit network
$S_l = \{1, 2, \cdots, \|S_l\|\}$	Set of stations on line l
$Q_l = \{1, 2, \cdots, \|Q_l\|\}$	Set of train services for line l
$T = \{T_1, T_2, \cdots, T_K\}$	Set of time periods
l	Index of lines, $l \in L$
$s,\ s'$	Index of stations, $s \in S_l$
$q,\ q'$	Index of train services, $q \in Q_l$
C_0	Loading capacity of each train
H_s	Set of transfer feeder lines with respect to station s
T_k	Index of the kth time period, $k = 1, 2, \ldots, K$
T_k^*	The ending time of period T_k
δ	Duration of each time period
$R_s(T_k)$	Arrival rate of passengers at station s during the time period T_k
t_{min}^W	Minimum allowable waiting time for transfer passengers
t_{max}^W	Maximum allowable waiting time for transfer passengers
t_{ideal}^W	Ideal waiting time for transfer passengers
R_{qs}^A	Percentage of passengers alighting from train q at station s
$R_{s's}^r$	Percentage of passengers transferring from station s' to station s
$t_{s's}^T$	Transfer walking time from station s' to station s
t_s^r	Running time between station s and the previous station $s-1$
$t_{q,l}/\bar{t}_{q,l}$	Lower/upper bound of the arrival time of the first train for line l
h_{min}/h_{max}	Minimum/maximum headway of two adjacent trains
d_{min}/d_{max}	Minimum/maximum train dwell times at stations

Table 3
List of decision variables.

Variable	Definition
t_{qs}^D	Departure time of train $q \in Q_l$ from station $s \in S_l$ on line $l \in L$
t_{qs}^A	Arrival time of train $q \in Q_l$ at station $s \in S_l$ on line $l \in L$
t_{qs}^d	Dwell time of train $q \in Q_l$ at station $s \in S_l$ on line $l \in L$
P_{qs}^B	Number of passengers who board on train q successfully at station s
P_{qs}^A	Number of passengers who alight from train q upon arriving at station s
P_{qs}^O	Number of in-vehicle passengers after train q departing from station s
P_{qs}^R	Number of passengers who remain waiting at station s after train q departs
P_{qs}^E	Number of passengers who enter station s between times $t_{(q-1)s}^D$ and t_{qs}^D
P_{qs}^T	Number of passengers who transfer from other lines and arrive at station s between times $t_{(q-1)s}^D$ and t_{qs}^D
C_{qs}	Remaining capacity of train q after passengers alight at station s

where $t_{q,l}$ and $\bar{t}_{q,l}$ represent the earliest and latest departure times for the first train of line l at the first station, respectively.

According to Assumption 1, the traveling time t_s^r of the trains at each section is assumed to be a predetermined value. Thus, the arrival and departure times of the trains at each station are limited by the arrival time at the origin station and the dwell times at the subsequent stations. Constraints (2) and constraints (3) strictly restrict the relationship between the arrival and departure times of the trains at each station.

$$t_{qs}^A - t_{q(s-1)}^D = t_s^r, \forall s \in S_l \backslash \{1\}, q \in Q_l, l \in L \tag{2}$$

$$t_{qs}^D - t_{qs}^A - t_{qs}^d = 0, \forall s \in S_l, q \in Q_l, l \in L \tag{3}$$

To ensure operational safety of the urban rail transit network, a set of safety and operational constraints (4) and (5) are formulated to impose the train following headway and dwelling time at each station, which are given by

$$h_{\min} \leqslant t_{qs}^D - t_{(q-1)s}^D \leqslant h_{\max}, \forall s \in S_l, q \in Q_l \backslash \{1\}, l \in L \tag{4}$$

$$d_{\min} \leqslant t_{qs}^d \leqslant d_{\max}, \forall s \in S_l, q \in Q_l, l \in L. \tag{5}$$

In constraint (4), h_{max} and h_{min} respectively denote the maximum and minimum headway times, specified by the signaling system. In constraint (5), d_{max} and d_{min} denote the maximum and minimum dwell times of trains at each station, respectively.

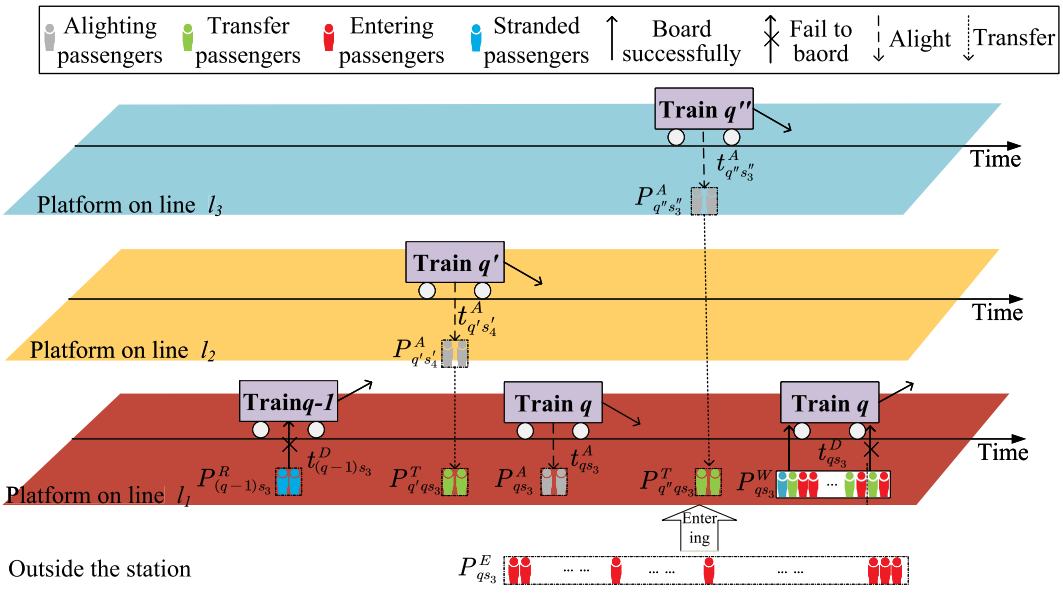


Fig. 4. An illustration of passengers at a transfer station. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2.3. Passenger flow constraints

The modeling of passenger flow in an urban rail network is very complex because of the transfer of passengers (Liu and Ceder, 2018). Specifically, at a transfer station, there are three types of passengers waiting on the platform during the departure of two adjacent trains, as depicted in Fig. 4. Type 1 comprises stranded passengers (represented by the blue color in Fig. 4) who fail to board on the previous trains when the capacity of the previous trains arriving at the station is less than the volume of passengers waiting on the platform. Type 2 refers to the passengers entering the platform from outside the station during this time period (represented by red color in Fig. 4). Type 3 is the passengers who transfer to this line from the other lines (represented by the green color in Fig. 4). Next, We modeled these three types of passenger flow, respectively.

Passenger type 1: Because of the limited capacity of trains, when the number of waiting passengers exceeds the remaining capacity of the arriving train q, some of the waiting passengers will be stranded at the platform, that is, they cannot board the incoming train. Constraint (6) imposes the relationship between the number of stranded passengers, waiting passengers, and boarding passengers.

$$P_{qs}^{R} = P_{qs}^{W} - P_{qs}^{B}, \quad \forall s \in S_l, q \in Q_l, l \in L \tag{6}$$

where P_{qs}^{R} denotes the number of passengers stranded by train q; P_{qs}^{W} denotes the total number of passengers who wait for train q at station s; and P_{qs}^{B} denotes the number of passengers who can successfully board train q at station s.

Passenger type 2: The number of passengers who enter the platform from outside the station within the departure interval can be expressed as constraint (7), that is, the red crowd in Fig. 4.

$$P_{qs}^{E} = (T^{*}(t_{(q-1)s}^{D}) - t_{(q-1)s}^{D})R_s(t_{(q-1)s}^{D}) + (t_{qs}^{D} - T^{*}(t_{(q-1)s}^{D}))R_s(t_{qs}^{D}), \quad \forall s \in S_l, q \in Q_l \backslash \{1\}, l \in L \tag{7}$$

$$P_{qs}^{E} = R_s(t_{qs}^{D}) \cdot t_s^{e}, \quad \forall s \in S_l, q = 1, l \in L \tag{8}$$

where $R_s(t_{qs}^{D})$ denotes the passenger arrival rate of the time period corresponding to the time that train q departs from s, and $T^{*}(t_{(q-1)s}^{D})$ denotes the ending time of the time period, in which train $q-1$ departs from station s. For the first train $q = 1$ on each line l, the value of P_{qs}^{E} can be simply calculated using constraint (8), where t_s^{e} is a given parameter that indicates the time allowed for passengers to enter the station s before the first train departs from station s. Here, we note that constraint (7) is a nonlinear constraint that cannot be handled directly. We will further reformulate the constraints into linear inequalities in Section 3.2.

Passenger type 3: Since there might be several elements in set H_s, that is, transfer station s is connected to multiple lines, the total number of transfer passengers to train q is the sum of all transfer passengers from different feeder lines (that is, all the green crowds in Fig. 4), as expressed in constraint (9).

$$P_{qs}^{T} = \sum_{l' \in H_s} \sum_{q' \in Q_{l'}} P_{q'qs}^{T}, \quad \forall s \in S_l, q \in Q_l, l \in L \tag{9}$$

where $P^T_{q'qs}$ denotes the number of transfer passengers from train q' to train q. Note that $P^T_{q'qs}$ is limited by the following constraints (10).

$$\sum_{q\in Q_l} P^T_{q'qs} = P^A_{q's} \cdot R^r_{s's}, \forall\{(s',s)\in S^r | s\in S_l, s'\in S_{l'}\}, q'\in Q_{l'}, l, l'\in L. \tag{10}$$

Constraint (10) ensures that the sum of the number of transfer passengers from train q' to all trains on the aim line is equal to the number of alighting passengers of q' at the transfer station multiplied by the transfer rate, where $P^A_{q's}$ represents the number of passengers alighting from train q' and $R^r_{s's}$ represents the transfer rate of from station s' to station s with $(s',s)\in S^r$.

Next, we consider modeling the constraints with respect to the boarding and alighting processes of the passengers. First, we denote constraints (11),

$$P^A_{qs} = P^O_{q(s-1)} \cdot R^A_{qs}, \quad \forall s\in S_l\backslash\{1\}, q\in Q_l, l\in L \tag{11}$$

which ensures that the number of passengers getting off the train q, that is, P^A_{qs}, is equal to the number of passengers in-vehicle before arriving at station s (i.e., $P^O_{q(s-1)}$) multiplied by the alighting rate R^A_{qs}. For the first station of each line l, we have

$$P^A_{qs} = 0, \quad \forall s = 1, q\in Q_l, l\in L. \tag{12}$$

Under constraint (11), the number of in-vehicle passengers is limited by the following equality:

$$P^O_{qs} = P^O_{q(s-1)} - P^A_{qs} + P^B_{qs}, \quad \forall s\in S_l\backslash\{1\}, q\in Q_l, l\in L, \tag{13}$$

This indicates that the number of in-vehicle passengers equals to the number of passengers in the previous station minus the number of alighting passengers added to the number of boarding passengers. In constraint (13), we have

$$P^B_{qs} = \min\left\{P^W_{qs}, C_{qs}\right\}, \quad \forall s\in S_l\backslash\{1\}, q\in Q_l, l\in L \tag{14}$$

indicating that the number of boarding passengers is the smaller value among the number of passengers waiting at the station (i.e., P^W_{qs}) and the remaining capacity C_{qs} of the train at this station. The nonlinear constraints (14) can be equivalently replaced by the following linear constraints:

$$P^B_{qs} \leqslant P^W_{qs}, \quad \forall s\in S_l\backslash\{1\}, q\in Q_l, l\in L \tag{15}$$

$$P^B_{qs} \leqslant C_{qs}, \quad \forall s\in S_l\backslash\{1\}, q\in Q_l, l\in L \tag{16}$$

In the above constraints, the calculation of C_{qs} is expressed by constraints (17) and (18):

$$C_{qs} = C_0, \quad \forall s = 1, q\in Q_l, l\in L \tag{17}$$

$$C_{qs} = C_0 - P^O_{q(s-1)} + P^A_{qs}, \forall s\in S_l\backslash\{1\}, q\in Q_l, l\in L, \tag{18}$$

This implies that the remaining capacity of the train at the first station is equal to the full capacity, whereas at other stations, it is equal to the full capacity minus the in-vehicle passengers.

Finally, we impose constraints (19) and (20) to limit the flow of waiting at each station.

$$P^W_{qs} = P^E_{qs} + P^T_{qs} + P^R_{(q-1)s}, \quad \forall s\in S_l, q\in Q_l\backslash\{1\}, l\in L, \tag{19}$$

$$P^W_{qs} = P^E_{qs} + P^T_{qs}, \quad \forall s\in S_l, q = 1, l\in L, \tag{20}$$

where P^W_{qs} is the total number of passengers waiting for train q at station s, expressed as the sum of the three types of passengers defined above. Note that for the first train of each line, there are only passengers of types 2 and 3, as given in constraint (20).

2.2.4. Objective function

Next, we respectively introduce the two objectives into our mathematical formulation.

Objective 1: Passenger waiting time

As expressed in several studies (Barrena et al., 2014; Samà et al., 2015), passenger waiting time is a key indicator for evaluating the service quality of urban rail lines. Because we consider three types of passengers in an urban rail network,we next present the waiting time of each passenger type within each time interval $[t^D_{(q-1)s}, t^D_{qs}]$.

$$W^R_{qs} = (t^D_{qs} - t^D_{(q-1)s}) \cdot P^R_{(q-1)s}, \quad \forall s\in S_l, q\in Q_l\backslash\{1\}, l\in L \tag{21}$$

$$W^E_{qs} = \int_{x=t^D_{(q-1)s}}^{t^D_{qs}} \int_{y=0}^{x-t^D_{(q-1)s}} R_s(x)dydx, \quad \forall s\in S_l, q\in Q_l\backslash\{1\}, l\in L \tag{22}$$

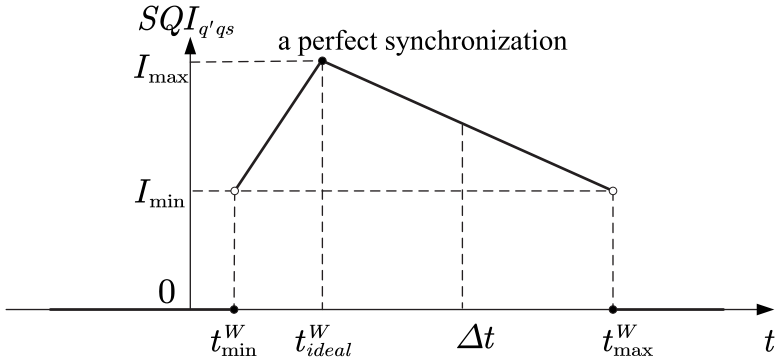


Fig. 5. Relationship between SQI and waiting time.

$$W^T_{qs} = \sum_{l' \in H_s} \sum_{q' \in Q_{l'}} (t^D_{qs} - t^A_{q's'} - t^T_{s's}) \cdot P^T_{q'qs}, \quad \forall s \in S_l, q \in Q_l, l \in L \tag{23}$$

Eqs. (21), (22) and (23) respectively denote the waiting times of passengers of type 1, 2, and 3, respectively, which are calculated by multiplying the waiting time of each passenger by the number of passengers.

Remark 2.1. Note that in (22), it is assumed that $R_s(x)$ is a continuous function (i.e., passenger arriving rate) with respect to time x. As we assumed that the passenger demand in each period T_k is constant, the calculation of passenger waiting time in (22) can be approximated by

$$\begin{aligned} W^E_{qs} =& \frac{1}{2} R_s(t^D_{(q-1)s}) \cdot (T^*(t^D_{(q-1)s}) - t^D_{(q-1)s})^2 \\ &+ \frac{1}{2}(t^D_{qs} - T^*(t^D_{(q-1)s}))[2R_s(t^D_{(q-1)s}) \cdot (T^*(t^D_{(q-1)s}) - t^D_{(q-1)s}) \\ &+ R_s(t^D_{qs}) \cdot (t^D_{qs} - T^*(t^D_{(q-1)s}))], \quad \forall s \in S_l, q \in Q_l \backslash \{1\}, l \in L. \end{aligned} \tag{24}$$

In particular, for the first train $q = 1$ on each line, because there is no previous train to inquire, the waiting time of entering passengers can be calculated approximately as follows:

$$W^E_{qs} = \frac{1}{2} R_s(t^D_{qs}) \cdot (t^e_s)^2, \quad \forall s \in S_l, q = 1, l \in L \tag{25}$$

Then, we obtain one of the objective functions, that is, the minimization of the passenger waiting time in the transit network, as expressed in Eq. (26).

$$\min W = \sum_{l \in L} \sum_{q \in Q_l} \sum_{s \in S_l} W_{qs} = \sum_{l \in L} \sum_{q \in Q_l} \sum_{s \in S_l} \left(W^R_{qs} + W^E_{qs} + W^T_{qs} \right). \tag{26}$$

Objective 2: Synchronization quality index

For transfer stations, it may be inappropriate to minimize the transfer waiting time of passengers, because such a goal may cause some passengers (especially passengers with limited mobility) to be unable to complete the transfer and thus wait for a longer time. Because "just miss" is always disappointing and leads to complaints, it is often more comfortable for passengers to have extra waiting time after arriving at the platform of the aim line (Kwan and Chang, 2008; Wong et al., 2008). Therefore, we consider a timetable that allows passengers to make smooth transfers with a short waiting time as a well-designed timetable.

To balance the passenger waiting time and train synchronization quality, we define an ideal waiting time t^W_{ideal}, minimum waiting time t^W_{min}, and maximum waiting time t^W_{max} for the transfer passengers. Then, the value of SQI is defined as follows: when the waiting time of transfer passengers (termed as Δt) at the platform is equal to t^W_{ideal}, we regard the synchronization quality of the feeder train and aim train as "perfect", and set their SQI to the maximum value, i.e. equal to I_{max}, as shown in Fig. 5. When Δt is less than t^W_{min} or greater than t^W_{max}, two trains are considered to be poorly synchronized, which may cause some passengers to transfer unsuccessfully or cause complaints, and we set their SQI to 0. When the waiting time is in the interval $[t^W_{min}, t^W_{max}]$, passengers can board the train as long as there is a non-zero remaining train capacity, and we define the two trains as *connected*.

In summary, the calculation for $SQI_{q'qs}$ of two trains q' and q on different lines at the transfer station s is given by

$$SQI_{q'qs} = \begin{cases} [I_{min} + r_1(\Delta t - t^W_{min})], & \Delta t \in (t^W_{min}, t^W_{ideal}] \\ [I_{max} + r_2(\Delta t - t^W_{ideal})], & \Delta t \in [t^W_{ideal}, t^W_{max}) \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

where $r_1 = (I_{max} - I_{min})/(t^W_{ideal} - t^W_{min})$, $r_2 = (I_{min} - I_{max})/(t^W_{max} - t^W_{ideal})$, $\Delta t = t^D_{qs} - t^A_{q's'} - t^T_{s's}$, and s' denotes the transfer station of train q'. Note that $SQI_{q'qs}$ is a piecewise linear nonconvex function with respect to passenger waiting time Δt, and its equivalent reformulation is described in Section 3.1.

Considering that different transfer stations have different importance to a transit network because of their location, we introduce parameter $\omega_{s's}$ to measure its weight. Then, the second objective function, that is, the maximization of the SQI of trains, can be expressed as follows:

$$\max SQI = \sum_{l\in L}\sum_{l'\in L}\sum_{q\in Q_l}\sum_{q'\in Q_{l'}}\sum_{(s',s)\in S^r : s\in S_l, s'\in S_{l'}} w_{s's}\cdot SQI_{q'qs} \tag{28}$$

Note that these bi-objective functions, W and SQI, are in different ranges. We adopt a weighted sum method and introduce the normalization factors W_{max}, W_{min}, SQI_{max}, and SQI_{min} and weighted parameters ω_1, ω_2 (see e.g., Yin et al. (2022)). Finally, the bi-objective functions can be denoted by a trade-off objective function, as expressed in Eq. (29).

$$\min J = \min\left(\omega_1 \cdot \frac{W - W_{min}}{W_{max} - W_{min}} - \omega_2 \cdot \frac{SQI - SQI_{min}}{SQI_{max} - SQI_{min}}\right) \tag{29}$$

where the values of ω_1 and ω_2 represent the importance of passenger waiting time W and transfer convenience SQI in the network, respectively.

After defining of constraints and objective functions, our MIP model for the train timetable synchronization problem is given by

$$\begin{cases} \text{obj:} & \min J \\ \text{s.t.} & \text{constraints (1)–(13)} \\ & \text{constraints (15)–(20).} \end{cases} \tag{30}$$

3. Model reformulation and mathematical properties

3.1. Reformulation of SQI

In the MIP formulation (30), the value of $SQI_{q'qs}$ defined in (27) is a piecewise linear function with nonconvex and semi-continuous properties, which means that even if it is a univariate function, solving this optimization problem can still be NP-hard (Vielma, 2015). Therefore, in this section, we reformulate SQI by introducing several sets of new variables and analyze the mathematical properties of the new formulations.

Our reformulation, termed as BCC, is based on the basic convex combination of vertices in a piecewise linear function. Specifically, we denote continuous variables $\lambda^i_{q'qs}$, $i \in \{1, 2, \cdots, 7\}$ and binary variables $y^j_{q'qs}$, $j \in \{1, \cdots, 4\}$ for each $SQI_{q'qs}$. Then, we can obtain the following equivalent reformulation with respect to $SQI_{q'qs}$. In (31), T^A and T^B denote the minimum and maximum values of Δt, respectively.

$$\text{(BCC)}\quad \begin{cases} T^A\lambda^1_{q'qs} + t^W_{min}(\lambda^2_{q'qs} + \lambda^3_{q'qs}) + t^W_{ideal}\lambda^4_{q'qs} + t^W_{max}(\lambda^5_{q'qs} + \lambda^6_{q'qs}) + T^B\lambda^7_{q'qs} = \Delta t \\ SQI_{q'qs} = 0\lambda^1_{q'qs} + 0\lambda^2_{q'qs} + I_{min}\lambda^3_{q'qs} + I_{max}\lambda^4_{q'qs} + I_{min}\lambda^5_{q'qs} + 0\lambda^6_{q'qs} + 0\lambda^7_{q'qs} \\ \lambda^1_{q'qs} + \lambda^2_{q'qs} = y^1_{q'qs} \\ \lambda^3_{q'qs} + \lambda^4_{q'qs} = y^2_{q'qs} \\ \lambda^4_{q'qs} + \lambda^5_{q'qs} = y^3_{q'qs} \\ \lambda^6_{q'qs} + \lambda^7_{q'qs} = y^4_{q'qs} \\ y^1_{q'qs} + y^2_{q'qs} + y^3_{q'qs} + y^4_{q'qs} = 1 \\ \lambda^i_{q'qs} \geqslant 0, \forall i \in \{1, \cdots, 7\} \\ y^1_{q'qs}, y^2_{q'qs}, y^3_{q'qs}, y^4_{q'qs} \in \{0, 1\} \end{cases} \tag{31}$$

Using the BBC reformulation in (31), we can then replace the piecewise function $SQI_{q'qs}$ by a series of new variables $\lambda^i_{q'qs}$ and $y^j_{q'qs}$ and constraints with linear forms, which can be handled by commercial solvers. For detailed explanation on the BCC reformulation, we present an example in Appendix A.

Next, we considered the tightness of the aforementioned reformulations. For simplicity, we denote the piecewise linear function as f, and the point on the function as (x, z), that is, $f(x) = z$. In other words, x and z represent Δt and $SQI_{q'qs}$ respectively. Theoretically, the tightness of MIP formulations can be quantified using *sharp* and *ideal*. Specifically, a set *gr(f)* in MIP is *sharp* if and only if the projection onto the x and z variables of its LP relaxation is exactly *conv(gr(f))*, as indicated in Vielma (2015). Additionally, an MIP formulation is denoted as *locally ideal* if its LP relaxation has basic feasible solutions that automatically satisfy the integral requirements of the x variables (Padberg and Rijal, 1996; Padberg, 2000).

Property 3.1. *The formulation BCC is both* sharp *and* locally ideal.

Proof. See Appendix A.

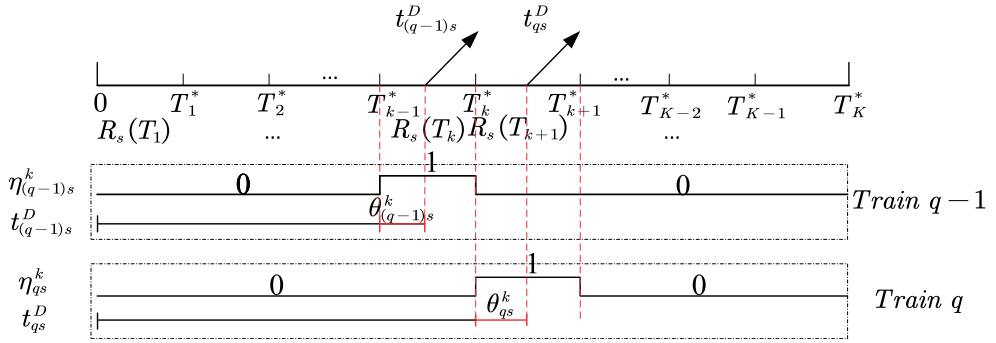


Fig. 6. An example of feasible values of η^k_{qs} and θ^k_{qs}.

3.2. Reformulation of passenger flow constraints

Next, we show how to linearize constraint (7) by introducing additional binary variables that are used to indicate the departure time of trains. Specifically, we denote the binary variable $\eta^k_{qs} = 1$ when train q departs from station s in the time interval T_k, otherwise, $\eta^k_{qs} = 0$. Then, we denote the continuous variable θ^k_{qs}, which represents the offset of the departure time of train q from station s with respect to T_k, as illustrated in Fig. 6. Here, parameter T^*_k represents the ending time of period T_k, where $T^*_0 = 0$. Thus, the relation between t^D_{qs}, η^k_{qs} and θ^k_{qs} can be expressed as

$$t^D_{qs} = \sum_{k=1}^{K} T^*_{k-1} \cdot \eta^k_{qs} + \theta^k_{qs}, \forall s \in S_l, q \in Q_l, l \in L \tag{32}$$

$$\sum_{k=1}^{K} \eta^k_{qs} = 1, \forall s \in S_l, q \in Q_l, l \in L \tag{33}$$

$$0 \leqslant \theta^k_{qs} \leqslant (T^*_k - T^*_{k-1}) \cdot \eta^k_{qs}, \forall k \in \{1, 2, \cdots, K\}, s \in S_l, q \in Q_l, l \in L \tag{34}$$

where constraint (32) guarantees that the departure time t^D_{qs} is uniquely determined by η^k_{qs} and θ^k_{qs}, constraint (33) imposes that the departure time of a train belongs to exactly one interval, and constraint (34) imposes the bounds on variable θ^k_{qs}.

To express the number of passengers entering the departure time interval linearly, we also need to introduce the binary variable b^k_{qs}, which is 1 if train q departs from station s in a period $j \leqslant k$, and b^k_{qs} can express as

$$b^k_{qs} = \sum_{j=1}^{k} \eta^j_{qs}, \forall k \in \{1, 2, \cdots, K\}, s \in S_l, q \in Q_l, l \in L. \tag{35}$$

The reformulation procedure is similar to that in Niu et al. (2015), in which cumulative binary variables were adopted to indicate train departure. Using these variables and inequalities, the linear representation of constraint (7) is given by:

$$P^E_{qs} = \sum_{k=1}^{K} R_s(T_k)(T^*_k - T^*_{k-1})(b^k_{(q-1)s} - b^k_{qs}) - \sum_{k=1}^{K} R_s(T_k)\theta^k_{(q-1)s} + \sum_{k=1}^{K} R_s(T_k)\theta^k_{qs}, \forall s \in S_l, q \in Q_l \backslash \{1\}, l \in L. \tag{36}$$

Then, we handle the nonlinearity of the passenger waiting time in Eq. (24), which is a cubic function, not supported by the commercial solvers. After introducing new variables η^k_{qs} and θ^k_{qs}, Eq. (24) can be transformed into the quadratic function given in (37):

$$\begin{aligned} W^E_{qs} = \sum_{k=1}^{K} R_s(T_k)\{ \frac{1}{2} \cdot [(\delta\eta^k_{(q-1)s} - \theta^k_{(q-1)s})^2 + (\theta^k_{qs} - (\eta^k_{(q-1)s} - \Delta b^k_{qs})\delta)^2] + \theta^{k+1}_{qs}(\delta\eta^k_{(q-1)s} - \theta^k_{(q-1)s}) \\ - (\delta\eta^k_{qs} - \theta^k_{qs})(\delta\eta^k_{(q-1)s} - \theta^k_{(q-1)s})\}, \forall s \in S_l, q \in Q_l \backslash \{1\}, l \in L \end{aligned} \tag{37}$$

where $\Delta b^k_{qs} = b^k_{(q-1)s} - b^k_{qs}$ is used to indicate whether two trains $q-1$ and q leave the station in the same time period T_k, if so, Δb^k_{qs} equals 0, otherwise, it equals 1, and θ^{K+1}_{qs} is set to 0.

Table 4
Number of decision variables with respect to input parameters.

Variables	Number of variables	Number of the variables in the sample case
t^D_{qs}	$\sum_{l=1}^{\|L\|} \|Q_l\|$	$2+3=5$
t^d_{qs}	$\sum_{l=1}^{\|L\|} \|Q_l\| * \|S_l\|$	$2\times 3+3\times 3=15$
α	$\sum_{l=1}^{\|L\|} \sum_{s=1}^{\|Q_l\|} \sum_{H_{ls}} \|Q_l\| * \|Q_{l'}\| * 7$	$2\times 3\times 7=42$
y	$\sum_{l=1}^{\|L\|} \sum_{s=1}^{\|Q_l\|} \sum_{H_{ls}} \|Q_l\| * \|Q_{l'}\| * 3$	$2\times 3\times 3=18$
η	$\|T\| * (\sum_{l=1}^{\|L\|} \|Q_l\| + \sum_{l=1}^{\|L\|} \|Q_l\| * \|S_l\|)$	$4\times(2+3+2\times 3+3\times 3)=80$
θ	$\|T\| * (\sum_{l=1}^{\|L\|} \|Q_l\| + \sum_{l=1}^{\|L\|} \|Q_l\| * \|S_l\|)$	$4\times(2+3+2\times 3+3\times 3)=80$
b	$\|T\| * (\sum_{l=1}^{\|L\|} \|Q_l\| + \sum_{l=1}^{\|L\|} \|Q_l\| * \|S_l\|)$	$4\times(2+3+2\times 3+3\times 3)=80$

Our reformulated MIP model is given by

$$\begin{cases} \text{obj:} & \min J \\ \text{s.t.} & \text{constraints (1)–(6)} \\ & \text{constraints (8)–(13)} \\ & \text{constraints (15)–(20)} \\ & \text{constraints (31)–(36).} \end{cases}$$

Note that the reformulated MIP model has a quadratic objective function (associated with the waiting time of passengers) and linear constraints, which can be solved to the optimum by commercial solvers such as CPLEX and Gurobi (see e.g., Bliek et al., 2014), which can thus provide useful benchmarks for subsequent solution methodologies.

4. Solution methodologies

Although the MIP model in the previous section can be solved exactly by commercial solvers such as CPLEX and Gurobi, the computational time can be very long owing to the large number of integral variables. As indicated in Cai and Gho (1994) and Caprara et al. (2002), heuristic algorithms are effective approaches for solving the timetable scheduling problem, which has been proven to be an NP-hard problem. Nevertheless, a series of heuristics, such as genetic algorithms, simulated annealing algorithms, and particle swarm algorithms, easily fall into local optimal solutions. In this section, we first analyze the complexity of the proposed MIP model, and we design a hybrid adaptive large-scale neighbor search (HALNS) algorithm for large networks, to escape from local optimal solutions.

4.1. Complexity analysis

The complexity of the model originates from the scale of the metro network and the number of variables introduced. Obviously, the former is closely related to the number of lines $|L|$ in the network, the number of stations $|S_l|$ on each line l, the number of planned train services $|Q_l|$ running on each line l, and the number of time intervals $|T|$. The complexity of the latter mainly depends on the following five types of variables: λ and $\mathbf{y}$ that introduced to linearize the SQI, and $\boldsymbol{\eta}$, θ and $\mathbf{b}$ that introduced to linearize the constraints related to passenger flows. In addition, the first two sets of variables are related to the number of transfer stations in the network, and we note that $\boldsymbol{\eta}$, $\mathbf{b}$, and $\mathbf{y}$ are all integer variables, which makes it more difficult to solve the model.

Here, we use a specific small example to quantify the complexity of the analysis model, as shown in Table 4. Suppose a simple network is composed of two lines, A and B, and each line has three stations. Consider that two trains run on Line A and three trains run on Line B. There is only one transfer relation, i.e. station 2 of line A can be transferred to station 2 of line B. Thus the number of key variables t^D_{qs} and t^d_{qs} are 5 and 15, respectively, and the numbers of λ and y are 5 and 15, respectively. If the planned time domain is divided into four time intervals, then the numbers of η, θ and b are all 80. The number of variables increases exponentially with the size of the network, especially when considering that our model constraints will produce many intermediate variables, which means that it is difficult to solve the model using a commercial solver within an acceptable computational time.

4.2. Development of hybrid ALNS

ALNS is an efficient method for solving large-scale discrete combinational optimization problems (Dong et al., 2020; Canca et al., 2019). In particular, ALNS can select the search neighborhood according to the dynamic weight of the operators, and the neighbor solutions are generated through the destroy-and-repair process of ALNS. Currently, ALNS has been successfully applied to large-scale discrete optimization problems, such as vehicle routing/scheduling problems (Azi et al., 2014; Sacramento et al., 2019; Dayarian et al., 2016; Özarik et al., 2021), pickup and delivery problems (Li et al., 2016), and railway yard planning (Ruf and Cordeau, 2021) problems, etc.

In our study, we combined the advantages of ALNS with the traditional simulated annealing (SA) algorithm. In particular, we develop destroy-and-repair operators in ALNS and propose an acceptance probability function, motivated by SA, to accept good solutions with varying probabilities to limit the probability of falling into a local optimum.

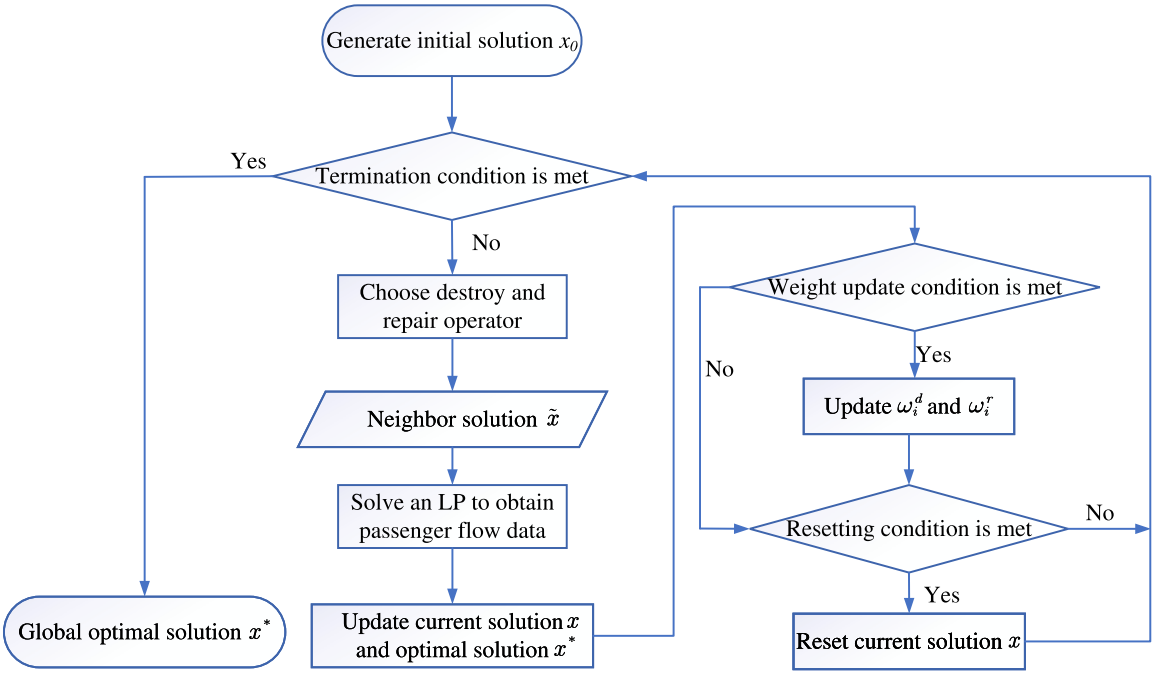


Fig. 7. Overall flow chart of HALNS.

4.2.1. Overall framework

The steps and flowchart of the proposed HALNS algorithm are presented in Algorithm 1 and Fig. 7. In our framework, we let the vectors $\boldsymbol{x}_0$, $\boldsymbol{x}$, $\boldsymbol{x}^*$, and $\tilde{\boldsymbol{x}}$ denote the initial, current, best, and neighbor solutions, respectively. Note that each solution vector $\boldsymbol{x}$ is composed of decision variables t_{qs}^D and t_{qs}^d, for $q \in Q_l$, $s \in S_l$, $l \in L$, i.e., the departure and dwelling times of all the trains.

Our initial solution $\boldsymbol{x}_0$ is a periodic timetable with a constant time headway (e.g., 2 min). In the iterations, we generate a set of neighbor solutions from the current solution $\boldsymbol{x}$, and then compute the objective function J with respect to the neighbor solution $\tilde{\boldsymbol{x}}$. Note that after determining decision variable $\tilde{\boldsymbol{x}}$, the remaining problem is a linear programming model that can be solved very efficiently to obtain the objective function value (see Yin et al., 2021). We then update the scores of the destroy/repair operators according to the value of the objective function. In step 5 of HALNS, the weight updating formula is given by

$$w_i^d = (1-\sigma)w_i^d + \frac{\sigma\pi_i^d}{v_i^d} \tag{38}$$

$$w_i^r = (1-\sigma)w_i^r + \frac{\sigma\pi_i^r}{v_i^r} \tag{39}$$

where w_i^d / w_i^r and π_i^d / π_i^r represent the weight and score of the destroy/repair operator i, respectively, and v_i^d / v_i^r represents the number of times the destroy/repair operator i is used. After the neighbor solutions are generated, we use a simulated annealing process with a varying probabilities to determine whether a solution can be accepted. In particular, we note that we set the current solution $\boldsymbol{x}$ using a periodic timetable with a headway of random values generated within the minimum and maximum headway if no neighbor solution is accepted. This process is repeated until the termination condition is satisfied.

4.2.2. Destroy operator

In our HALNS algorithm, a neighbor solution is generated after two destroy-and-repair processes, which are of great significance to the effect of the algorithm. As described above, the critical decision variables in this study were the departure time t_{qs}^D of each train at the first station and the dwell time t_{qs}^d at every station. Thus, the destroy operator *destroy_i* is defined by selecting a random number N_i^D of variables from t_{qs}^D, removing these variables to a list $\mathbb{L}_D$, and then selecting a random number N_i^d of variables from t_{qs}^d and removing these variables to a list $\mathbb{L}^d$ to be repaired. In particular, in each iteration, we adopt four destroy operators in parallel, with different numbers of selected variables.

For example, as illustrated in Fig. 8, we choose a total of seven variables when N_i^D counts 2 and N_i^d counts 5 respectively in *destroy_*1, and a total of 10 variables when N_i^D counts 3 and N_i^d counts 7 in *destroy_*2. The values of N_i^D and N_i^d increase from *destroy_*1 to *destroy_*4, so that the degree of damage to the current solution gradually increases.

Algorithm 1 The procedure HALNS

Input: Input network topology data and dynamic passenger demand.
Step 1. Generate initial solution $\boldsymbol{x}_0$, $\boldsymbol{x} \leftarrow \boldsymbol{x}_0$, $\boldsymbol{x}^* \leftarrow \boldsymbol{x}_0$.
Step 2. Whether the termination condition is met:
  Step 2.1 If yes, output $\boldsymbol{x}^*$;
  Step 2.2 Otherwise, go to Step 3;
Step 3. Generate $\tilde{\boldsymbol{x}}$ according to ω.
Step 4. Update $\boldsymbol{x}^*$,$\boldsymbol{x}$ and $\boldsymbol{\pi}$:
  If $(f(\tilde{\boldsymbol{x}}) < f(\boldsymbol{x}^*))$,$\boldsymbol{x}^* \leftarrow \tilde{\boldsymbol{x}}$, $\boldsymbol{x} \leftarrow \tilde{\boldsymbol{x}}$, $\pi_i^d \leftarrow \pi_i^d + c_1$, $\pi_i^r \leftarrow \pi_i^r + c_1$;
  else if $(f(\tilde{\boldsymbol{x}}) < f(\boldsymbol{x}))$, $\boldsymbol{x} \leftarrow \tilde{\boldsymbol{x}}$, $\pi_i^d \leftarrow \pi_i^d + c_2$, $\pi_i^r \leftarrow \pi_i^r + c_2$;
  else $\boldsymbol{x} \leftarrow \tilde{\boldsymbol{x}}$ (with a certain probability), $\pi_i^d \leftarrow \pi_i^d + c_3$, $\pi_i^r \leftarrow \pi_i^r + c_3$;
Step 5. Whether the weight updating condition is met:
  Step 5.1 If yes, update weights by Eqs. (38);
  Step 5.2 Otherwise, go to Step 6;
Step 6. Whether the resetting condition is met:
  Step 6.1 If yes, reset $\boldsymbol{x}$;
  Step 6.2 Otherwise, go to Step 2;

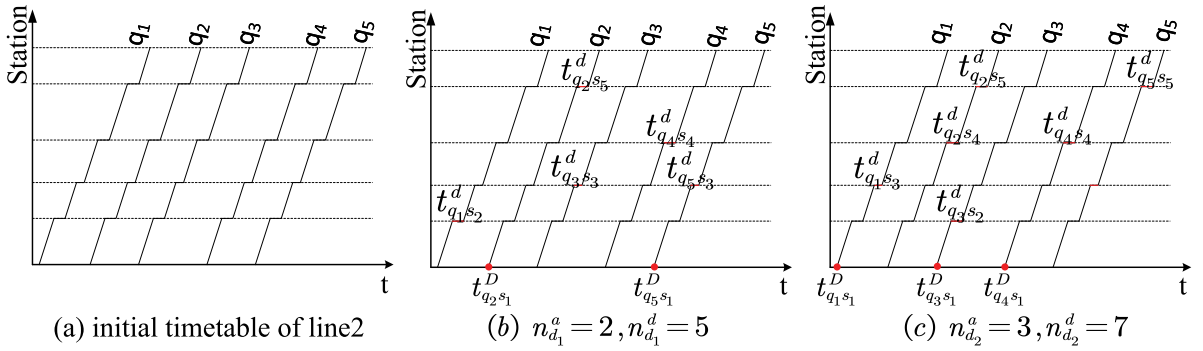


Fig. 8. Illustration of destroy operators.

4.2.3. Repair operator

Repair operators are very important for ALNS because they rebuild the destroyed solution, resulting in a new set of candidate solutions. In our problem, because the candidate solutions after the destroy operations are possibly infeasible, we need to repair them into feasible solutions. Our repair operator comprises two steps. In step one, we first repair the solutions with respect to t^D_{qs}, that is, the departure time of trains from the first station of each line. In particular, we begin with the trains in the destroyed solutions. Let $t^D_{q_u s}$ denote the first element of $\mathbb{L}_D$. According to constraints (4), we have

$$h_{min} \le t^D_{q_u s} - t^D_{q_{u-1} s} \le h_{max},$$

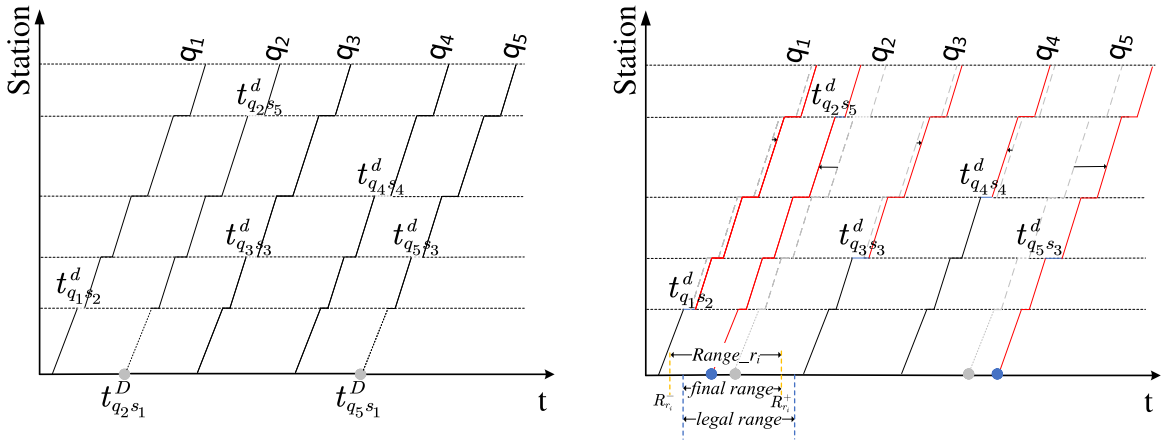
which derives a feasible range of variables $t^D_{q_u s}$, that is, $t^D_{q_u s} \in [h_{min} + t_{q_{u-1} s}, h_{max} + t^D_{q_{u-1} s}]$. We can then randomly select a value in its feasible range as the value of $t^D_{q_u s}$ in the neighbor solution. The following elements in list $\mathbb{L}_D$ are successively repaired in a similar manner, until set $\mathbb{L}_D$ becomes empty.

After determining all departure times, we can repair the dwell times at each station in $\mathbb{L}_d$. This process is similar to that of $\mathbb{L}_D$. We randomly chose a value from the feasible range $[d_{min}, d_{max}]$. Notably, after the dwell time is determined, all arrival and departure times after the station must be recalculated.

Similar to the destroy operators, we denote four repair operators in each iteration. Considering the destroyed timetable generated by *destroy*_1 in Fig. 8 as an example, we repair the departure time variables in the repair list according to the order of the train services. As shown in Fig. 9, we first need to repair $t^D_{q_2 s_1}$. According to the departure times of q_1 and q_3, we can determine a legal range that satisfies the safety headway constraints, and the intersection of this legal range and the custom range in the repair operator (i.e., $Range\_r_i$, $R^-_{r_i}$, and $R^+_{r_i}$, respectively, representing the left and right boundaries of this range) as the final range of variable $t^D_{q_2 s_1}$. We randomly select a value in the final range as the value of $t^D_{q_2 s_1}$ in the neighbor solution, and $t^D_{q_5 s_1}$ is obtained using the same process.

4.2.4. Acceptance and stopping criteria

By employing multiple destroy-and-repair operators, we can generate a set of candidate neighbor solutions for each iteration. To decide whether to accept a solution, we adopt an SA scheme that not only accepts improved solutions but also accepts degraded



(a) Timetable after destroy

(b) Timetable after repair

Fig. 9. Illustration of repair operators.

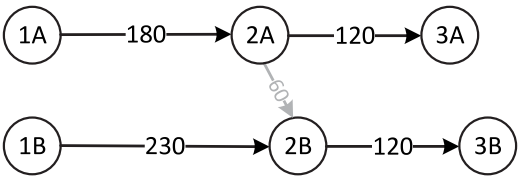


Fig. 10. Topology of the small-scale case study.

solutions with probabilities, which enables HALNS to escape from the local optimum. Specifically, we let $\mathbf{x}'$ denote one of the neighbor solutions of $\mathbf{x}$, and let $J(\mathbf{x}')$ and $J(\mathbf{x})$ denote the objective function values of $\mathbf{x}'$ and $\mathbf{x}$, respectively. If the neighbor solution $\mathbf{x}'$ is not as good as $\mathbf{x}$, that is, $J(\mathbf{x}') \geq J(\mathbf{x})$, we define a probability $e^{\frac{-J(\mathbf{x}')-J(\mathbf{x})}{TM}}$, which represents the acceptance probability of the neighbor solution $\mathbf{x}'$. Here, the acceptance probability decreases with the profit disparity and with parameter TM, that is, the temperature of the SA. The value of TM is set as follows: In the beginning of HALNS, TM is set to T_{init}. The value of TM is reduced after each iteration by multiplying TM by the cooling factor $\phi \in (0, 1)$.

We set two termination conditions for HALNS. First, we terminate the algorithm if the program running time reaches a set threshold. The second stopping criterion for HALNS is when the entire process reaches the maximum number of iterations.

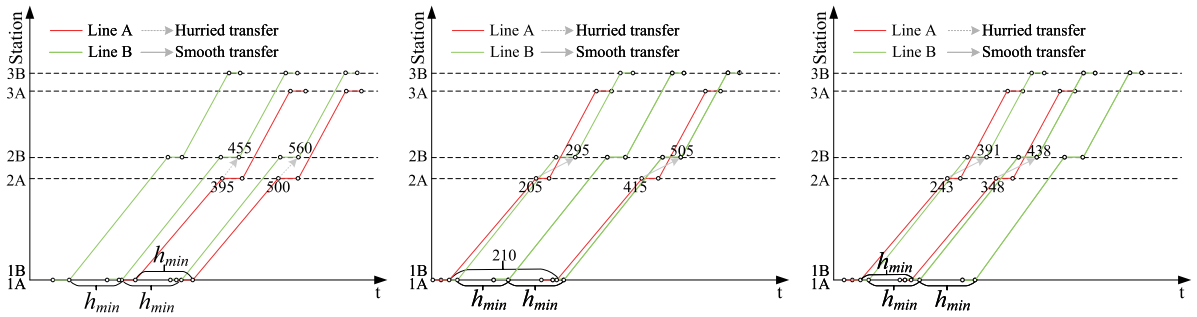
5. Numerical experiments

In this section, we describe two sets of experiments, including a series of small-scale cases and a series of real-world cases, to illustrate the effectiveness of our approach. All the experiments were implemented with C# language on a 64-bit computer with 16 GB RAM and an Intel I7-8700k CPU. The version of CPLEX solver is 12.10 Academic version.

5.1. Small-scale case study

Small-scale case studies considered only two lines with six stations, as shown in Fig. 10 (i.e., the same example as described in Section 4.1). Two trains run from stations 1A to 3A on line A, and three trains run from 1B to 3B on line B. The running times between the stations are marked with numbers on the horizontal line. Stations 2A and 2B are the same physical transfer stations, and passengers can transfer from 2A to 2B with a fixed walking time of 60 s. The other parameters associated with this small network are presented in Table 5. We divide the time horizon T into four periods, i.e., $T = \{T_1, T_2, T_3, T_4\}$. The time period was 600 s. The arrival, alighting, and transfer rates of passengers are listed in Table 6.

Next, we present two experimental groups. In the first group, we used CPLEX to solve our MIP model directly, with different objective functions, that is, maximizing SQI, minimizing W or minimizing J, in order to quantify the trade-off between different objectives. In the second group, we compared the computational performance of the three solution approaches, that is, CPLEX, our HALNS, and a standard genetic algorithm (GA), a widely used meta-heuristic algorithm. The GA process is detailed in Appendix B.



(a) Timetable with minimal passenger waiting time (b) Timetable with maximal SQI (c) Timetable with two objectives

Fig. 11. Train timetables with different objective functions.

Table 5
Parameter settings in the small-scale experiments.

Parameters	Value	Parameters	Value
C_0	600	t^W_{ideal}	30 s
t^W_{min}	0 s	t^W_{max}	90 s
I_{min}	1	I_{max}	2
h_{min}	105 s	h_{max}	600 s
d_{min}	25 s	d_{max}	45 s
$\bar{t}_{q,l}$	0 s	$\underline{t}_{q,l}$	360 s

Table 6
Passenger demand in the small case.

Station		1A	2A	3A	1B	2B	3B
Arrival rate (Person/second)	T_1	1.304	1.241	0	1.039	1.007	0
	T_2	1.734	1.673	0	1.402	1.563	0
	T_3	1.593	1.457	0	1.222	1.436	0
	T_4	1.304	1.341	0	1.139	1.222	0
Alighting rate		0	0.4	1	0	0.7	1
Transfer rate		$2A \rightarrow 2B$: 0.5					

5.1.1. Comparison of timetables with different objectives

We compare and analyze the timetables with the three different objective functions, that is, the minimization of passenger waiting time W, the maximization of SQI, and their trade-off. Because the network scale is small, we use CPLEX to solve the MIP model, which takes less than 3 s, and the results are presented in Fig. 11.

The first timetable shown in Fig. 11(a) is generated with minimized W, that is, the minimum passenger waiting time. We see that, the trains on each line run with a minimum interval, which undoubtedly minimizes the waiting time for passengers. However, the quality of synchronization between these two lines is very poor at this time because the waiting time of transfer passengers at the transfer station is also minimized to 0. This means that it is a hurried transfer; in other words, the passengers will witness the departure of the train as soon as they arrive at the platform of their aim line.

The second timetable, as shown in Fig. 11(b), is generated to maximize SQI. In contrast to the first timetable, shown in Fig. 11(a), the trains on line A no longer run at the minimum interval because the passenger waiting time is no longer the objective to be considered. Meanwhile, the synchronization between the two lines is optimal, and transfer passengers can board the train with an ideal waiting time. The third timetable shown in Fig. 11(c) considers both the transfer time and transfer convenience between these two lines. We can see that the transfer passengers can experience a smooth transfer while the waiting time of passengers in the network is minimized. Overall, we can clearly see straightforwardly that the trade-off between these two objectives W and SQI, instead of a single objective, is particularly important for the service quality of traveling passengers in an urban rail network.

5.1.2. Comparison among different solution approaches

Next, we compare the performances of the different approaches for solving the constructed models. In particular, we use these three methods on three instances, in which the network scale (number of lines, stations and trains) is gradually increased. For CPLEX, the time limit was set to 7200 s, and for GA/HALNS, the time limit was set to 600 s. The computational results are presented in Table 7. The first four columns in Table 7 present the number of lines, stations, train services, and connected station pairs in the metro network.

From the results in Table 7, we can see that CPLEX performs the best in the smallest instance, that is, in the case of two lines. CPLEX can obtain the global optimal solutions regardless of the objective function, although it requires as long as 2 hours when the

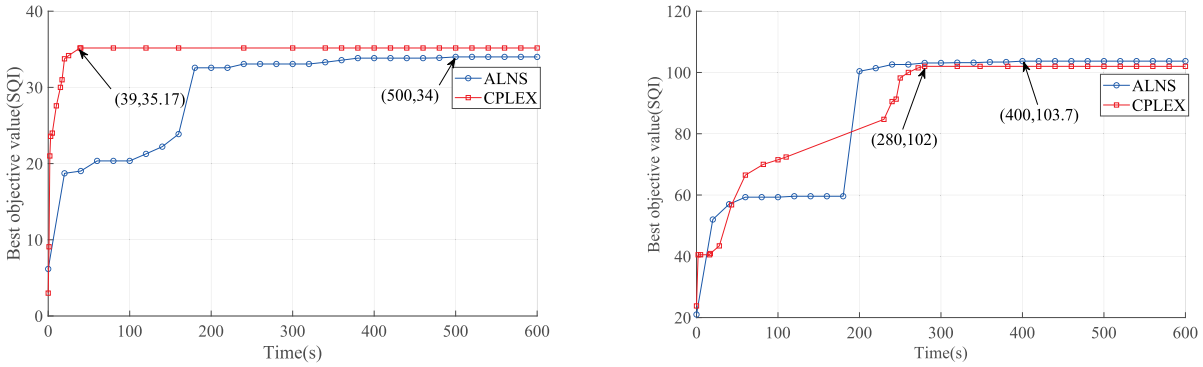


Fig. 12. Convergence tendency of the objective function SQI.

Table 7
Performance comparison of CPLEX, GA and HALNS.

Scale of network				Objective[a]	Best known objective value[b]	Solution method	Objective value	Computation time unit (s)	Opt gap (%)[c]
No.of line	No.of station	No.of train	No.of con						
2	6	5	1	SQI	4	CPLEX	**4**	0.739	0.0
						GA	2	600	100.0
						HALNS	**4**	0.033	0.0
				W	44 850	CPLEX	**44 850**	7200	0.0
						GA	52 962	600	15.3
						HALNS	45 036	600	0.5
				J	−0.969	CPLEX	**−0.969**	7200	0.0
						GA	−0.089	600	988.7
						HALNS	−0.869	600	11.5
3	42	18	4	SQI	35.2	CPLEX	**35.2**	322	0.0
						GA	16	600	120.0
						HALNS	34	600	3.5
				W	–	CPLEX	NFS[d]	–	–
						GA	535 579	600	–
						HALNS	**445 474**	600	–
				J	–	CPLEX	NFS	–	–
						GA	0.480	600	–
						HALNS	**0.434**	600	–
5	52	50	8	SQI	112	CPLEX	102	7200	8.9
						GA	64	600	75.0
						HALNS	**104**	600	7.7
				W	–	CPLEX	NFS	–	–
						GA	5 159 346	600	–
						HALNS	**4 632 504**	600	–
				J	–	CPLEX	NFS	–	–
						GA	4.621	600	–
						HALNS	**0.327**	600	–

The gap is calculated according to formula: (Current value - Best-known value)/(Current value) $\times 100\%$.

[a] SQI: maximize SQI; W: minimize W; J: minimize $J = \omega_1 \cdot \frac{W - W_{min}}{W_{max} - W_{min}} - \omega_2 \cdot \frac{SQI - SQI_{min}}{SQI_{max} - SQI_{min}}$.

[b] The best-known objective function value is obtained by CPLEX with a 24-h computation.

[c] Symbol "-" means that the optimal solution is found.

[d] Symbol "NFS" means that no feasible solution is found.

objective functions are W and J. Meanwhile, we can see that HALNS can also obtain near-optimal solutions with a much shorter computational time. Among these three methods, the GA performs the worst, as it always generates poor objective functions. This is because GA falls into a local optimum that is far from the global optimum. As we increase the number of lines to 3, we see that CPLEX cannot even return feasible solutions associated with the objective functions of W and J because of the large number of

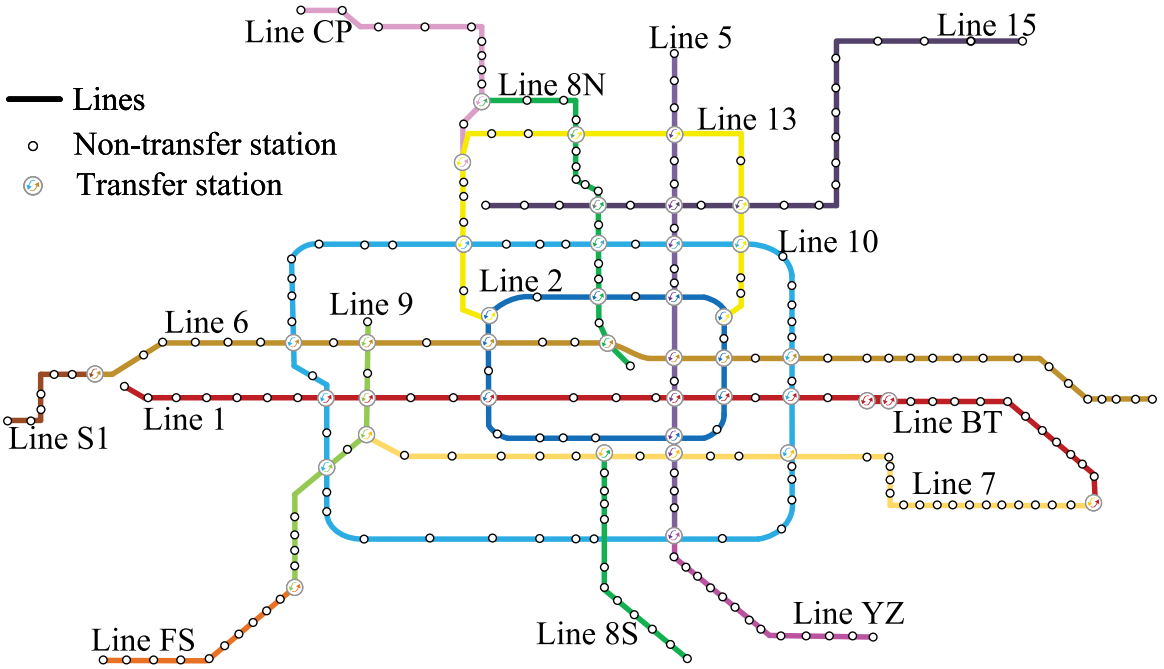
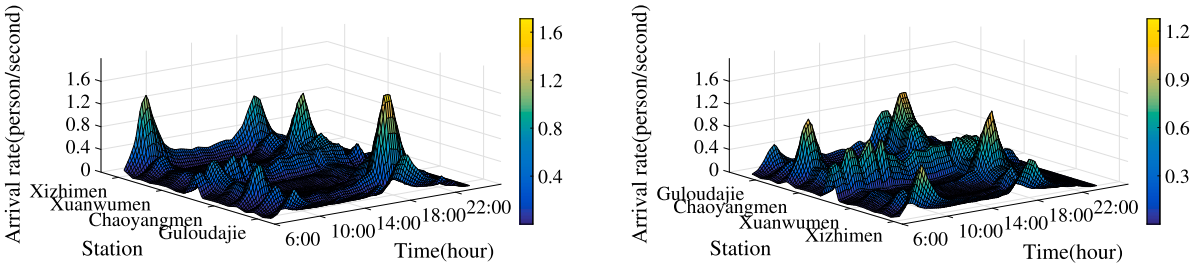


Fig. 13. Layout of Beijing metro network.



(a) Line 2 up direction

(b) Line 2 down direction

Fig. 14. Passenger arrival rate of Beijing Metro Line 2.

integer variables, whereas our HALNS can still generate high-quality solutions within the time limit. In particular, when the number of lines is further increased to 5, we see that HALNS evidently outperforms both CPLEX and GA for all the performance indicators. From these results, we conclude that CPLEX can be used only for very small-scale experiments, whereas our HALNS is more effective for solving medium or large-scale instances.

To better understand the performance of our HALNS method, we next show the convergence tendency of the objective function values using CPLEX and HALNS. Owing to the poor effect of the GA, we did not show the convergence trend of the GA. Fig. 12 presents the results for the case of three lines and five lines in Table 7. Note that because CPLEX cannot return feasible solutions with objective functions W and J, we only present the results of instances with objective function SQI. We can see from the results that CPLEX shows high efficiency at the beginning, but soon reaches a relatively stable value. From Fig. 12(b), we see that HALNS jumps out from the local optimum and obtains better solutions than CPLEX afterwards. The results reveal that our HALNS has better capabilities to generate high-quality solutions for large-scale instances, owing to its adaptive searching from the neighborhood solutions in the iterations.

5.2. Real-world case study

In this section, we conduct a series of large-scale experiments based on the real-world data from the Beijing metro network. As shown in Fig. 13, the considered Beijing metro network involves 16 physical lines and 158 stations, including 40 transfer stations. We use the passenger demand data from historical AFC data collected on January, 2020. For illustration purposes, Fig. 14 visualizes the passenger demand data in the upward and downward directions of Beijing metro line 2.

Table 8
Parameter settings of real-word scale experiments.

Time period	6:00–7:30	7:30-9:00	9:00–11:00	11:00–13:00	13:00–16:00	16:00–18:00	18:00–20:00	20:00–22:00
h_{min} (unit: s)	120	60	120	120	180	120	120	120
h_{max} (unit: s)	600	480	600	720	720	600	480	720
d_{min} (unit: s)	20	20	20	20	20	20	20	25
d_{max} (unit: s)	35	45	35	40	35	40	45	45

Table 9
Number of services of each line in the timetable used in Beijing metro.

Line	Num	Line	Num	Line	Num	Line	Num
1_up	278	7_up	174	10_up	294	CP_up	118
1_dn	258	7_dn	142	10_dn	291	CP_dn	220
2_up	238	8_up	176	13_up	159	YZ_up	161
2_dn	260	8_dn	226	13_dn	226	YZ_dn	137
5_up	278	8S_up	127	15_up	142	FS_up	151
5_dn	276	8S_dn	126	15_dn	166	FS_dn	197
6_up	243	9_up	277	S1_up	118	BT_up	255
6_dn	202	9_dn	279	S1_dn	119	BT_dn	204

Table 10
Performance comparison of HALNS with the existing timetable of the Beijing metro.

Instance		W(/p)	W1(/pt)	W2(/pt)	W3(/pt)	SQI	STP	PSP
1	OriT	413.5	250.8	270	153.7	34 455	21 543	0.435
	OptT	396.2	175.2	263.3	190.7	41 757	27 582	0.356
	Improvement	4.2%	30.1%	2.5%	6.5%↓[a]	21.2%	28.0%	18.3%
2	OriT	319.4	156.7	264.6	153.3	34 455	21 543	0.329
	OptT	309.9	101.2	254.6	181	40 873	27 050	0.248
	Improvement	3.0%	35.5%	3.8%	18.1%↓	18.6%	25.6%	24.8%
3	OriT	246	83.2	256.5	153	34 455	21 543	0.212
	OptT	245	45.7	240.2	171.2	40 454	26 707	0.136
	Improvement	0.4%	45.1%	6.3%	11.9%↓	17.4%	24.0%	35.9%
4	OriT	190.7	27.9	252.7	152.7	34 455	21 543	0.084
	OptT	189.8	27	251.4	147.5	38 144	25 345	0.082
	Improvement	0.4%	3.2%	0.5%	3.4%	10.7%	17.6%	2.5%
5	OriT	170.9	8.3	291.2	152.5	34 455	21 543	0.023
	OptT	170.3	8	286.2	142	38 317	25 444	0.022
	Improvement	0.4%	3.7%	1.7%	6.9%	11.2%	18.1%	2.5%
6	OriT	164.4	1.73	293.4	152.5	34 455	21 543	0.005
	OptT	163.7	1.67	285.3	147.5	37 801	24 925	0.005
	Improvement	0.4%	3.7%	2.8%	3.3%	9.7%	15.7%	0%
Average	Improvement	**1.5%**				**14.8%**		

OriT: Original timetable in Beijing metro; OptT: Optimized timetable.
STP: Number of successfully connected trains; PSP: Percentage of stranded passengers (unit: %).
[a]Symbol ↓ means that the performance is decreased.

The other experimental parameters, such as the maximum and minimum headway times, are listed in Table 8. The termination condition of all the experiments is that the computational time reaches 30 min. As revealed in the previous sections, neither CPLEX nor GA can return satisfactory results for large-scale instances. Thus, in the following section, we only adopt our HALNS to solve real-world instances.

5.2.1. Comparison with the existing timetable in beijing metro

In the first experiment, we compare our HALNS approach with the existing timetable in the Beijing metro for six instances (with historical data from six different days). These six instances were numbered from instance 1 to instance 6. For the Beijing metro, the operational time for each day ranges from 6 AM to 10 PM, with a certain number of services for each line (as presented in Table 9). In our HALNS, the numbers of services, that is, the values of $|Q_l|$ and $l \in L$ are set the same as those in the existing timetable.

Table 10 shows the performance improvement by our approach with respect to the existing timetable for the Beijing metro in detail. In Table 10, "$W(/p)$", "$W1(/p)$", "$W2(/p)$", and "$W3(/p)$" respectively represent the average waiting time for all the passengers, for type 1 passengers (i.e., stranded passengers), for type 2 passengers (i.e., newly arriving passengers), and for type 3 passengers (i.e., transfer passengers), respectively. "SQI", "STP" and "PSP" respectively represent the value of SQI indicators, the number of successfully *connected* trains (i.e., Δt is between t^W_{min} and t^W_{max}), and the percentage of stranded passengers at stations, respectively.

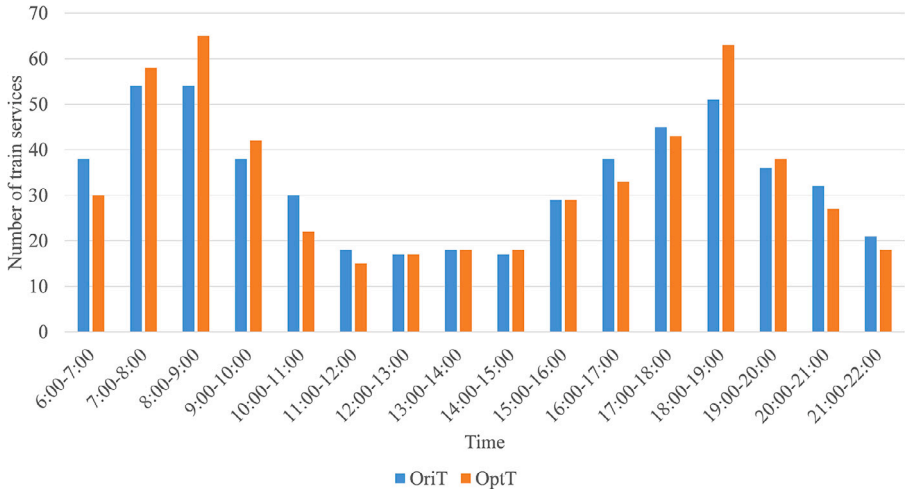


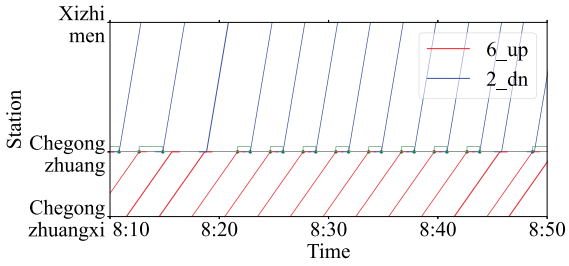
Fig. 15. Comparison of the number of services of line 1 before and after optimization.

From the results, we see that the optimized timetable has negligible effect on the passenger waiting time, as it achieves a slight improvement (within 4%) compared with the existing timetable in the Beijing metro. The reason is that the existing timetable in Beijing metro takes the waiting time of passengers as the main consideration. In contrast, we see that the optimized timetable remarkably improves the value of SQI by approximately 10% to 20% compared with the existing timetable of the Beijing metro. In particular, the number of successfully connected trains is significantly increased by more than 15%. This demonstrates the effectiveness of our approach for improving the synchronization of train timetables while reducing the passenger waiting time. An interesting phenomenon is that, although our model improves the values of SQI and STP, the average waiting time of transfer passengers increased only slightly in instances 1 and 3. A possible reason is that we consider the total waiting time of all passengers in the entire network as the optimization object rather than the waiting time of a certain type of passengers. In addition, from the last column of this table, we see that our optimized timetable can also reduce the number of stranded passengers. This is possibly due to fact that our synchronized timetables make the transfer of passengers smoother, thereby reducing the volume of waiting passengers from other lines and relieving the crowdedness of the transfer stations.

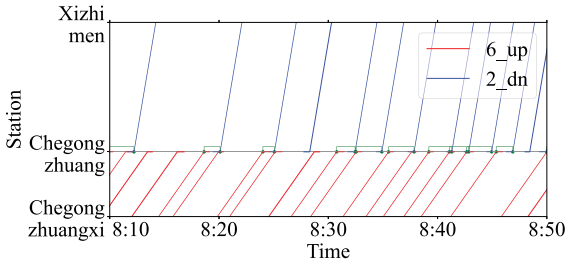
To better understand the difference between the optimized timetable and the existing timetable in the Beijing metro, we compared the number of train services during each hour before and after our optimization approach (i.e., OriT and OptT). The comparison results for line 1 are shown in Fig. 15 (the results of the remaining lines are given in Table 11 of Appendix C). The results show that, in general, our model allocates more services to the time periods when passenger demand is higher in the network. For the time periods when the passenger demand is relatively low, such as 11:00–12:00 or 20:00–22:00, our optimized timetable requires fewer trains than the existing timetable. This is consistent with the practical requirement that the service frequency should be adapted to the time-varying passenger demand.

Next, we used Beijing metro line 2 as an example to visualize the synchronized train timetable. In Beijing metro, line 2 is a loop line that covers the downtown area of Beijing. As one of the busiest lines in Beijing metro, line 2 is connected with line 6 at Chegongzhuang station and Chaoyangmen station, with line 5 at Chongwenmen station, and with line 1 at Jianguomen station. In Fig. 16, we present the original and optimized timetables at the four transfer stations. We can see that our optimized timetable increased the number of synchronized trains on the existing timetable of the Beijing metro. In addition, we see that the transfer convenience at the Chaoyangmen and Jianguomen stations in the existing timetable is quite poor, and our timetable synchronization approach can remarkably improve the transfer convenience at these two stations.

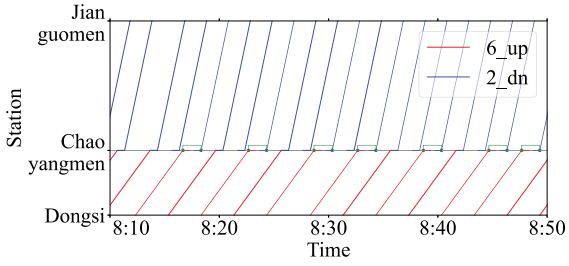
Finally, we aim to analyze the overall improvement in SQI at all transfer stations. To this end, we considered three different time periods: morning peak hours, off-peak hours and evening peak hours. In each period, we sum the values of SQI for all transfer stations and visualize the results in Fig. 17. Based on the results shown in Fig. 17, we can clearly see that our approach evidently improves the value of SQI compared to the existing timetable for all three time periods. There are two interesting phenomena observed in the results. First, our model has the most prominent effect during peak hours (that is, in Fig. 17(a) and Fig. 17(c)), whereas the improvement during off-peak hours (that is, in Fig. 17(b)) is not very obvious. A possible reason is that the capacity of trains during off-peak hours is sufficient to meet passenger demand, and there are fewer trains with lower frequencies. It is difficult to further improve the SQI during this period. Second, we can see that in some busy interchange stations (such as Fuxingmen, Dongdan, and Guomao), the optimized SQI is much higher (see e.g., Fig. 17(a)). This is because the passenger demand at these interchange stations is large, and the train services are more intensive. In this sense, the impact on the waiting time of passengers in the entire network was greater. From a practical viewpoint, our research results clearly indicate that the rail managers should pay more attention to the connectivity of trains at busy transfer stations, which can significantly improve the overall service quality of passengers.



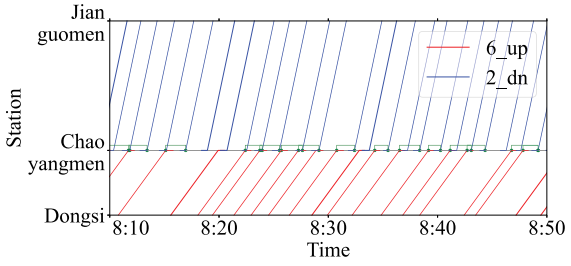
(a) Chegongzhuang station (OriT)



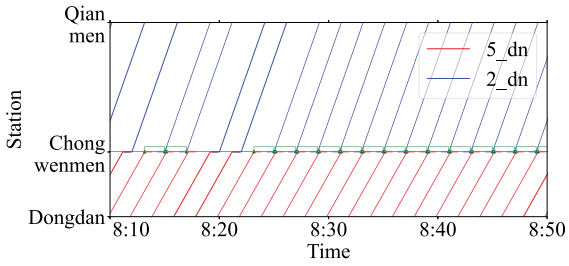
(b) Chegongzhuang station (OptT)



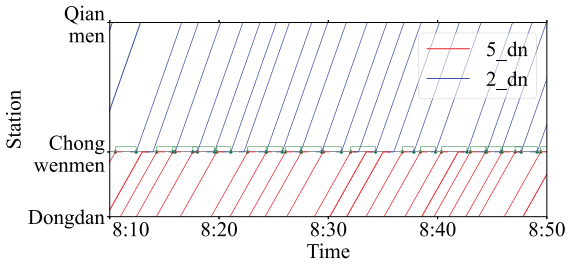
(c) Chaoyangmen station (OriT)



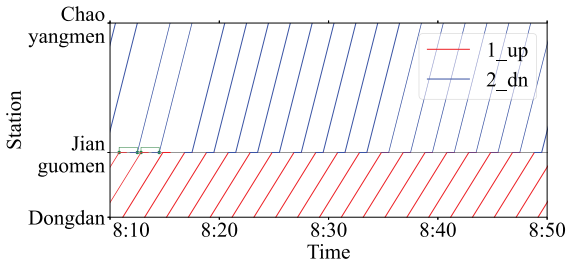
(d) Chaoyangmen station (OptT)



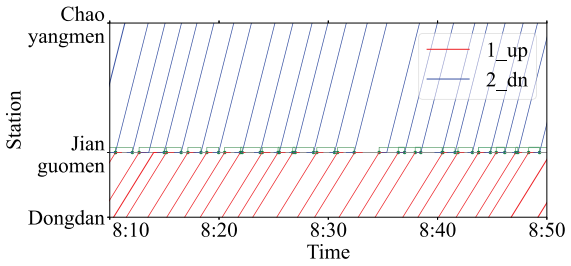
(e) Chongwenmen station (OriT)



(f) Chongwenmen station (OptT)



(g) Jianguomen station (OriT)

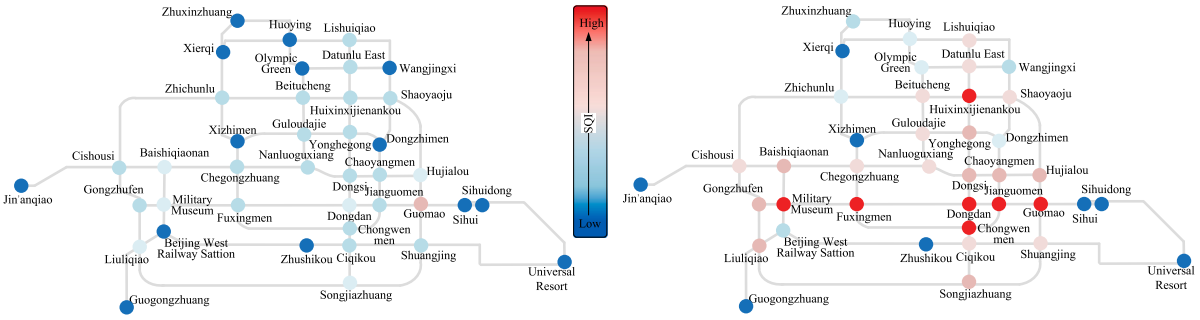


(h) Jianguomen station (OptT)

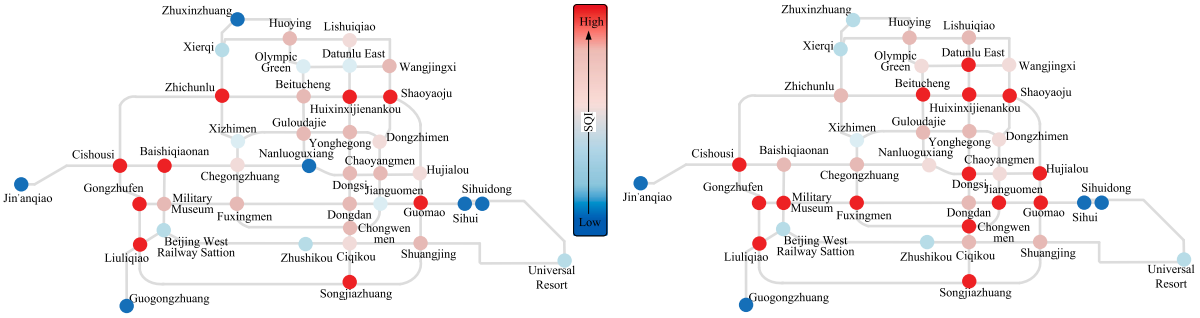
Fig. 16. Comparison of timetables at four transfer stations (left: original timetable; right: optimized timetable; green dots represent successfully connected trains).

5.2.2. Analysis of pareto optimal solutions

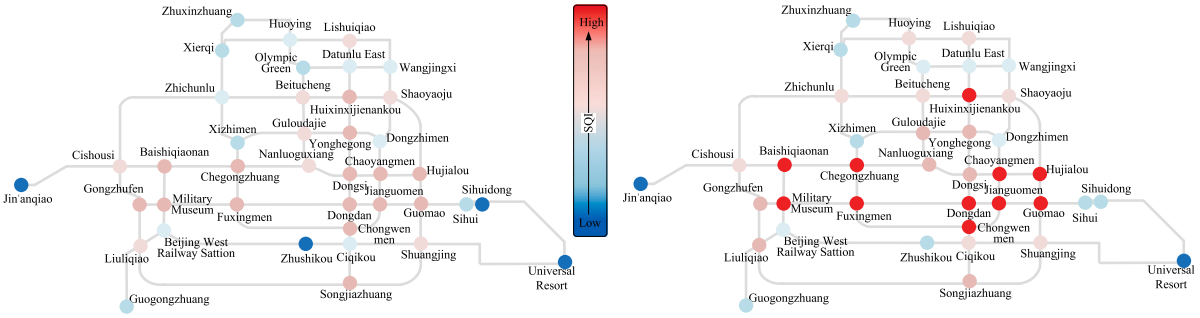
As revealed in the previous sections, the bi-objective function, that is, passenger waiting time W and synchronization quality SQI, has trade-off relations. There is an emerging requirement for rail managers to quantitatively analyze the relationship between these two objectives. To analyze the trade-off between the passenger waiting time W and transfer convenience SQI, we adopt a weight sum method, which is a widely used method in multi-objective optimization. In the weight sum method, we gradually



(a) Values of SQI during morning peak hours (left: OriT; right: OptT)



(b) Values of SQI during off-peak hours (left: OriT; right: OptT)



(c) Values of SQI during evening peak hours (left: OriT; right: OptT)

Fig. 17. Visualization of SQI values for different time periods (red color dot means better SQI). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

changed ω_1 from 0 to 1 (and correspondingly changed ω_2 from 1 to 0) to obtain the non-dominated solutions and to generate the Pareto frontiers for W and SQI. The results are presented in Fig. 18. The figure clear shows that when the proportion of SQI slowly increases from 0, the improvement in SQI is very evident, which highlights the benefits of considering transfer convenience in designing train timetables for the Beijing metro network. As we continue to increase the proportion of SQI after the two proportions become equal, the improvement in SQI becomes slower, which implies a strong trade-off relationship between W and SQI. The results indicate that the rail managers should carefully consider these two objectives by investigating the percentage of transfer passengers and designing a timetable that achieves the best trade-off performance.

6. Conclusion

In this study, we investigated the synchronized timetable optimization problem considering time-varying passengers in a complex urban rail transit network. We developed a bi-objective mathematical model that optimizes the waiting time of passengers and the synchronization quality of the timetable defined by an SQI function. We proposed a series of model reformulation techniques that

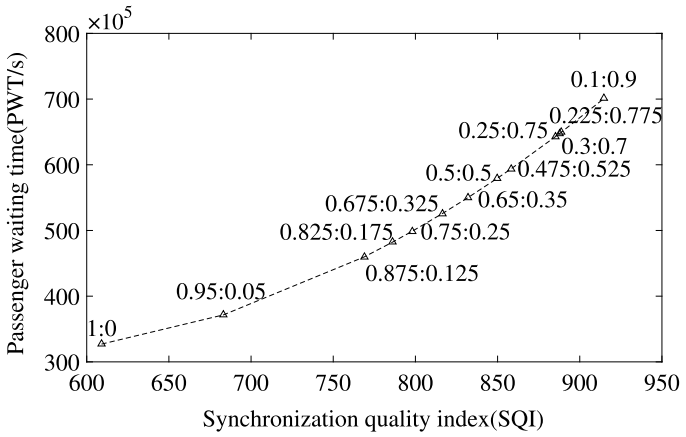


Fig. 18. Pareto frontier for W and SQI with different values of ω_1 and ω_2.

allow us to solve small-scale or medium-scale instances by MIP solvers exactly. To solve large-scale instances more efficiently, we further developed a HALNS algorithm by combining ALNS with simulated annealing.

To verify the effectiveness of our approach, we conducted two sets of experiments, involving a set of small-scale instances and a set of real-world instances based on the historical data from the Beijing metro network. The small-scale instances demonstrate that our HALNS algorithm outperforms the commercial solver CPLEX and metaheuristic GA. In real-world experiments, we compared the timetable of our approach with the existing timetable used in the Beijing metro. The main conclusions are as follows:

1. The two objectives, i.e., waiting time of passengers and synchronization quality SQI, have a strong trade-off relationship and the rail manager should carefully consider the importance of each objective when designing the network timetable.
2. Compared with the existing timetable in Beijing metro, the model proposed in this paper can reduce the waiting time of passengers in the whole network while optimizing the quality of timetable synchronization. In addition, our approach can also reduce the number of stranded passengers and alleviate the congestion at key stations.
3. In oversaturated situations, e.g., during peak-hours, optimization of a single objective (i.e., passenger waiting time) is limited, whereas the bi-objective optimization can be much more effective for improving the traveling experiences of passengers.

Future research will focus on the following aspects. (1) In urban rail networks, the transfer of passengers between connected lines is a very complex process because of the heterogeneous types of passengers, for example, disabled passengers or kids who have long transfer walking times. Future research will focus on the synchronization of train timetables considering these complicated factors. (2) Another interesting direction is to extend the framework into a bi-level approach by considering the choice of paths regarding transfer passengers. (3) Finally, the development of future urban mobility ecosystems requires a multimodal transportation network. Thus, the synchronization of timetables with different transport modes, for example, metro, transit, and sharing vehicles, is also be a very promising direction.

CRediT authorship contribution statement

Jiateng Yin: Study conception and design, Analysis and interpretation of results, Writing – original draft, Revision. **Miao Wang:** Data collection, Analysis and interpretation of results, Writing – original draft. **Andrea D'Ariano:** Writing – original draft, Revision. **Jinlei Zhang:** Study conception and design, Revision. **Lixing Yang:** Study conception and design.

Data availability

Data will be made available on request.

Acknowledgment

This study was supported by the National Natural Science Foundation of China (Nos. 72288101 and 71825004), by the Beijing Municipal Natural Science Foundation (No. 4222051), and by the State Key Laboratory of Rail Traffic Control and Safety (No. RCS2022ZZ003).

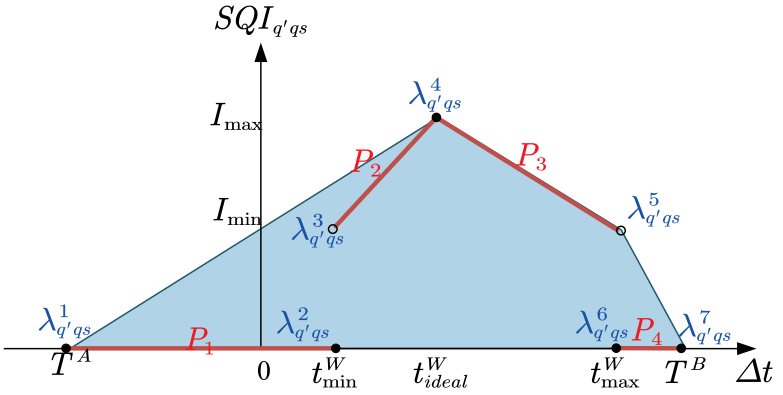


Fig. 19. Illustration of the reformulation of $SQI_{q'qs}$.

	individual 1	individual 2	individual 3	individual 4
gene a	0 200 400 600 800	20 200 380 530 700	50 210 390 550 710	35 205 380 540 700
gene b	10 200 390 580 770	0 200 400 610 780	0 180 360 540 720	20 230 440 640 840
gene c	0 180 380 570 770	10 190 380 580 800	30 200 370 540 710	20 200 380 560 740

Fig. 20. Sample chromosomes for the example in Fig. 2.

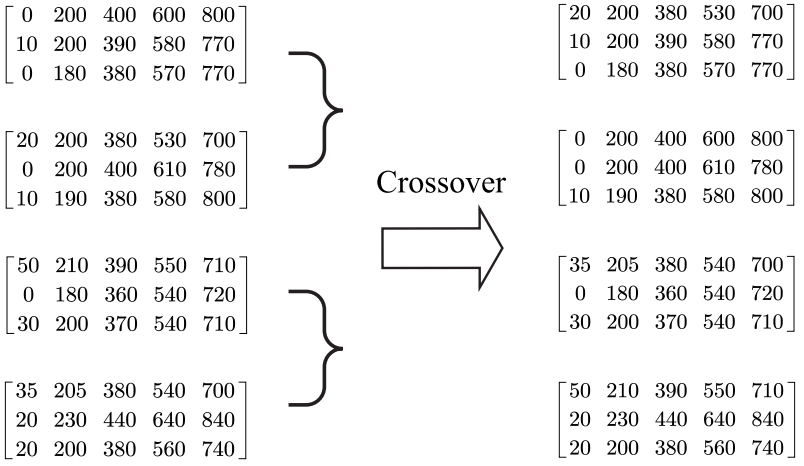


Fig. 21. Illustration of crossover operation for the sample chromosomes.

Appendix A. Explanation of BCC reformulation

We use Fig. 19 to help explain the BCC reformulation. First, we note that each $\lambda^i_{q'qs}$, $i \in \{1, 2, \cdots, 7\}$, in the convex combination formulations represents the corresponding vertex on the piecewise linear function. We can denote the coordinates of the ith vertex with (a_{xi}, a_{zi}), and the first two equations of BCC formulation indicate that any point on the function can be expressed as a convex combination of all vertices. In particular, when any $y^j_{q'qs}$ equals 1, it means that at this time Δt and the corresponding $SQI_{q'qs}$ belong to the polyhedron P_j.

Proof of Property 3.1. We first consider the sharpness of BCC formulation. Here, we use the Fourier–Motzkin elimination method to compute the projection of (31) onto the x and z variables, denoted by BCCxz. We can get the following formulation

$$\text{(BCCxz)} \quad \begin{cases} (t^W_{ideal} - T^A)z \le I_{max}x + I_{max}T^A \\ (t^W_{max} - t^W_{ideal})z \le (I_{min} - I_{max})x + I_{max}t^W_{max} - I_{min}t^W_{ideal} \\ (T^B - t^W_{max})z \le -I_{min}x - I_{min}T^B \\ T^A \le x \le T^B, z \ge 0 \end{cases} \tag{40}$$

Table 11
Number of services of each line in the timetable of Beijing metro.

Time		L1	L2	L5	L6	L7	L8	L8S	L9	L10	L13	L15	LS1	LCP	LYZ	LFS	LBT
6:00–7:00	OriT	38	22	35	26	18	23	20	35	42	21	26	17	18	21	25	32
	OptT	30	14	28	16	14	13	10	24	34	15	16	7	8	14	14	20
7:00–8:00	OriT	54	41	60	39	26	38	24	54	60	26	25	22	31	31	29	44
	OptT	58	41	62	45	22	40	28	61	65	21	27	24	37	36	37	48
8:00–9:00	OriT	54	57	60	38	28	42	24	57	60	34	23	21	32	29	36	48
	OptT	65	70	70	51	42	54	34	68	66	49	37	43	45	36	51	65
9:00–10:00	OriT	38	47	46	33	22	33	16	46	60	33	20	13	31	20	24	36
	OptT	42	49	53	33	23	35	14	44	63	33	21	6	31	21	18	33
10:00–11:00	OriT	30	32	24	24	17	19	12	26	48	21	16	12	18	12	16	20
	OptT	22	27	16	20	11	17	12	22	44	19	15	9	10	10	11	17
11:00–12:00	OriT	18	19	20	18	17	19	12	20	26	21	16	12	12	12	12	19
	OptT	15	17	18	14	15	15	9	20	23	19	10	7	13	9	9	15
12:00–13:00	OriT	17	18	20	17	17	16	10	20	20	18	15	12	12	12	12	16
	OptT	17	18	19	19	22	18	13	20	20	20	13	16	10	12	16	19
13:00–14:00	OriT	18	18	20	17	16	18	12	20	19	19	15	12	12	12	12	18
	OptT	18	18	21	18	15	16	12	21	20	19	20	11	13	13	11	17
14:00–15:00	OriT	17	18	21	18	18	16	12	21	2	19	14	12	12	12	12	16
	OptT	18	18	21	15	12	18	12	17	22	16	10	10	11	11	13	17
15:00–16:00	OriT	29	18	23	18	18	18	13	25	25	20	16	12	12	12	18	18
	OptT	29	17	23	17	21	13	9	27	22	20	16	13	11	10	14	15
16:00–17:00	OriT	38	22	36	34	20	20	18	46	36	20	18	13	12	19	23	26
	OptT	33	20	32	33	18	20	20	38	35	17	18	8	10	14	20	25
17:00–18:00	OriT	45	44	60	42	25	26	20	53	52	25	19	18	20	28	33	48
	OptT	43	46	62	43	35	27	19	55	54	23	14	24	19	36	32	49
18:00–19:00	OriT	51	57	54	48	28	40	21	52	54	33	22	18	32	29	33	42
	OptT	63	65	58	59	23	51	29	64	58	44	37	26	39	36	48	47
19:00–20:00	OriT	36	43	32	33	18	34	15	37	52	33	21	18	32	25	30	36
	OptT	38	40	32	29	20	32	13	36	54	33	20	12	36	23	28	39
20:00–21:00	OriT	32	22	24	27	17	20	12	24	9	20	21	13	31	12	19	22
	OptT	27	21	22	25	12	17	14	21	5	18	20	12	29	12	17	17
21:00–22:00	OriT	21	19	19	13	11	20	12	20	0	20	21	12	21	12	14	18
	OptT	18	17	17	8	11	16	5	18	0	19	14	9	16	5	9	16

which is exactly $conv(gr(f))$. Thus, the sharpness is proved.

Next, we consider the local ideal property of BCC formulation. For description convenience, we first define set $Y = \{0, 0, I_{min}, I_{max}, I_{min}, 0, 0\}$, i.e., each element of Y, denoted by γ_i, corresponds to the coefficient of variable $\lambda^i_{q'qs}$, $i = 1, 2, \cdots, 7$, for $SQI_{q'qs}$ in (31). Because the LP relaxation of BCC is convex, the optimal solution must be a certain vertex. For BCC, we assume that the optimal solution is vertex $\bar{v}$ (associated with vectors of variables $\bar{x}, \bar{z}, \bar{\lambda}$ and $\bar{y}$), such that $\bar{y}^j \in (0, 1), j \in \bar{J} \subseteq \{1, 2, 3, 4\}$ and $\sum_{j\in\bar{J}} \bar{y}^j = 1$. According to the constraints between variables $y^i_{q'qs}$ and $\lambda^i_{q'qs}$, we can also derive set $\bar{I} \subseteq \{1, 2, \dots, 7\}$, such that $\bar{\lambda}^i \in (0, 1)$, $i \in \bar{I}$. Let $q = argmax(\{\gamma_i | i \in \bar{I}\})$, i.e., $\gamma_q > \gamma_i$ for $\forall i \in \bar{I}\backslash\{q\}$. Then, we define a new solution $(\tilde{x}, \tilde{z}, \tilde{\lambda}, \tilde{y})$ that complies with (31) and $\tilde{\lambda}^q = 1$, which also implies that variable $\tilde{y}$ is integral, i.e., each element of $\tilde{y}$ takes value of 0 or 1. Thus, we have

$$\tilde{z} = \sum_{k\in\{1,2,\dots,7\}} \gamma_k \tilde{\lambda}^k = \gamma_q$$

and

$$\bar{z} = \sum_{i\in\bar{I}} \gamma_i \bar{\lambda}^i$$

Since $\sum_{i\in\bar{I}} \bar{\lambda}^i \leqslant 1$ and $\gamma_q > \gamma_i$ for $i \in \bar{I}\backslash\{q\}$, we obtain that $\tilde{z} > \bar{z}$, which contradicts $\bar{z}$ being the optimal solution value, and thus we prove the local ideal property of BCC. □

Appendix B. Implementation procedure of GA

In this appendix, we present the implementation details of GA in Section 5. The steps of the GA are given as algorithm 2. In our GA, after the parameters are input and the population size is determined, we create the first population randomly which includes N individuals. Take the network shown in Fig. 2 as an example. Assume there are 5 trains that run on each line. Some sample chromosomes "A" of individuals in first population are shown in Fig. 20. In Step 2, we assess the fitness of every individual and the

best one is denoted by x_k^*. If x_k^* is better than x^*, we update the best solution x^*. In Step 3, we crossover the best individual x_k^* with other individuals in the population (the crossover operation is depicted as Fig. 21). We combine gene "a" on chromosome "A" of individual 1 and genes "b" and "c" on chromosome "A" of individual 2 to form a new chromosome "A". In Step 4, we mutate the new individuals created by crossover and generate the new population X_{k+1}. Then we set the iteration time $k = k+1$. In Step 5, we judge whether the termination criteria of the algorithm is reached (the same as that of HALNS). If yes, we terminate the algorithm; otherwise, the algorithm goes back to step 2.

Algorithm 2 The procedure GA

Input: Input network topology data and dynamic passenger demand. Set the population size N.
Step 1. Create the first population $X_0 = \{x_0^1, x_0^2, \cdots, x_0^N\}$ randomly, $x^* \leftarrow x_0^1$, set $k = 0$.
Step 2. Assess the fitness of every individual.
  Step 2.1 If $x_k^* < x^*$, $x^* \leftarrow x_k^*$;
  Step 2.2 Otherwise, go to Step 3;
Step 3. Select the best individual x_k^* to crossover with other individual.
Step 4. Create the new population X_{k+1} by mutation, set $k = k+1$.
Step 5. Whether the termination condition is met:
  Step 5.1 If yes, output x^*;
  Step 5.2 Otherwise, go to Step 2;

Appendix C. Number of services of each line before and after optimization

See Table 11.

References

Azi, N., Gendreau, M., Potvin, J., 2014. An adaptive large neighborhood search for a vehicle routing problem with multiple routes. Comput. Oper. Res. 41, 167–173.
Barrena, E., Canca, D., Coelho, L., Laporte, G., 2014. Single-line rail rapid transit timetabling under dynamic passenger demand. Transp. Res. B 70, 134–150.
Bliek, C., Bonami, P., Lodi, A., 2014. Solving mixed-integer quadratic programming problems with IBM-CPLEX: a progress report. In: Proceedings of the Twenty-Sixth RAMP Symposium. Tokyo, pp. 171–180.
Cai, X., Gho, C., 1994. A fast heuristic for the train scheduling problem. Comput. Oper. Res. 21 (5), 499–510.
2021. China Association of Metros (CAMET) Annual Report. https://www.camet.org.cn/english.htm.
Canca, D., De-Los-Santos, A., Laporte, G., Mesa, J.A., 2019. Integrated railway rapid transit network design and line planning problem with maximum profit. Transp. Res. 127, 1–30.
Cao, Z., Ceder, A., Li, D., Zhang, S., 2019. Optimal synchronization and coordination of actual passenger-rail timetables. J. Intell. Transp. Syst. 23 (3), 231–249.
Caprara, A., Fischetti, M., Toth, P., 2002. Modeling and solving the train timetabling problem. Oper. Res. 50 (5), 851–861.
Ceder, A., Golany, B., Tal, O., 2001. Creating bus timetables with maximal synchronization. Transp. Res. A 35 (10), 913–928.
Chai, S., Yin, J., D'Ariano, A., Samà, M., Tang, T., 2023. Scheduling of coupled train platoons for metro networks: a passenger demand-oriented approach. Transp. Res. Rec. 2677 (2), 1671–1689.
Chu, J.C., Korsesthakarn, K., Hsu, Y.T., Wu, H.Y., 2019. Models and a solution algorithm for planning transfer synchronization of bus timetables. Transp. Res. 131, 247–266.
Dayarian, I., Crainic, T., Gendreau, M., Rei, W., 2016. An adaptive large-neighborhood search heuristic for a multi-period vehicle routing problem. Transp. Res. 95, 95–123.
Dong, X., Li, D., Yin, Y., Ding, S., Cao, Z., 2020. Integrated optimization of train stop planning and timetabling for commuter railways with an extended adaptive large neighborhood search metaheuristic approach. Transp. Res. C 117, 102681.
Guo, X., Sun, H., Wu, J., Jin, J., Gao, Z., 2017. Multiperiod-based timetable optimization for metro transit networks. Transp. Res. B 96, 46–67.
Guo, X., Wu, J., Sun, H., Liu, R., Gao, Z., 2016. Timetable coordination of first trains in urban railway network: a case study of beijing. Appl. Math. Model. 40 (17–18).
Hassannayebi, E., Zegordi, S.H., Amin-Naseri, M.R., Yaghini, M., 2016. Demand-oriented timetable design for urban rail transit under stochastic demand. Iran. Inst. Ind. Eng. 9 (3), 28–56.
Huang, Y., Ma Nn Ino, C., Yang, L., Tang, T., 2020. Coupling time-indexed and big-m formulations for real-time train scheduling during metro service disruptions. Transp. Res. B 133, 38–61.
Ibarra-Rojas, O.J., Rios-Solis, Y.A., 2012. Modeling and solving the train timetabling problem. Transp. Res. B 46, 599–614.
Kang, L., Zhu, X., 2016. Modeling and solving the first train timetabling problem with minimal missed trains in subway networks. Transp. Res. B 93, 17–36.
Kwan, C., Chang, C.S., 2008. Timetable synchronization of mass rapid transit system using multiobjective evolutionary approach. IEEE Trans. Syst. Man Cybern. C 38, 636–648.
Li, Y., Chen, H., Prins, C., 2016. Adaptive large neighborhood search for the pickup and delivery problem with time windows, profits, and reserved requests. European J. Oper. Res. 252, 27–38.
Li, X., Lu, L., Zheng, P., Huang, Z., 2019. Timetable coordination of the first trains for subway network with maximum passenger perceived transfer quality. IEEE Access 99, 1.
Liu, T., Ceder, A., 2018. Integrated public transport timetable synchronization and vehicle scheduling with demand assignment: A bi-objective bi-level model using deficit function approach. Transp. Res. B 117, 935–955.
Liu, R., Li, S., Yang, L., 2020. Collaborative optimization for metro train scheduling and train connections combined with passenger flow control strategy. Omega 90.
Liu, T., Oded, C., Konstantinos, G., 2021. A review of public transport transfer coordination at the tactical planning phase. Transp. Res. C 133, 103450.
Niu, H., Zhou, X., 2013. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. Transp. Res. C 36, 212–230.

Niu, H., Zhou, Z., Gao, R., 2015. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints. Transp. Res. B 76, 117–135.
Özarik, S.S., Veelenturf, L.P., Van Woensel, T., Laporte, G., 2021. Optimizing e-commerce last-mile vehicle routing and scheduling under uncertain customer presence. Transp. Res. E 148, 102263.
Padberg, M., 2000. Approximating separable nonlinear functions via mixed zero–one programs. Oper. Res. Lett. 27 (1), 1–5.
Padberg, M., Rijal, M.P., 1996. Location, Scheduling, Design, and Integer Programming. Springer.
Ruf, M., Cordeau, J., 2021. Adaptive large neighborhood search for integrated planning in railroad classification yards. Transp. Res. B 150, 26–51.
Sacramento, D., Pisinger, D., Ropke, S., 2019. An adaptive large neighborhood search metaheuristic for the vehicle routing problem with drones. Transp. Res. C 102, 289–315.
Samà, M., Meloni, C., D'Ariano, A., Corman, F., 2015. A multi-criteria decision support methodology for real-time train scheduling. J. Rail. Transport. Pla. 5 (3), 146–162.
Shafahi, Y., Khani, A., 2010. A practical model for transfer optimization in a transit network: model formulations and solutions. Transp. Res. A 44 (6), 377–389.
Shi, J., Yang, L., Yang, J., Gao, Z., 2018. Service-oriented train timetabling with collaborative passenger flow control on an oversaturated metro line: An integer linear optimization approach. Transp. Res. B 110, 26–59.
Sun, L., Jin, J., Lee, D., Axhausen, K., 2014. Demand-driven timetable design for metro services. Transp. Res. C 46, 284–299.
Tian, X., Niu, H., 2018. A bi-objective model with sequential search algorithm for optimizing network-wide train timetables. Comput. Ind. Eng. 127 (1), 1259–1272.
Vielma, J.P., 2015. Mixed integer linear programming formulation techniques. SIAM Rev. 57 (1), 3–57.
Wang, M., Chen, X., Yin, J., Su, S., D'Ariano, A., Wang, Y., Tang, T., 2021. A bi-objective optimization model for coordinated train timetabling in rail transit networks. In: IEEE International Intelligent Transportation Systems Conference. pp. 2350–2355.
Wang, Y., D'Ariano, A., Yin, J., Meng, L., Tang, T., Ning, B., 2018. Passenger demand oriented train scheduling and rolling stock circulation planning for an urban rail transit line. Transp. Res. B 118 (12), 193–227.
Wang, Y., Li, D., Cao, Z., 2020. Integrated timetable synchronization optimization with capacity constraint under time-dependent demand for a rail transit network. Comput. Ind. Eng. 142 (Apr.), 106374.
Wong, R., Yuen, T., Fung, K., Leung, J., 2008. Optimizing timetable synchronization for rail mass transit. Transp. Sci. 42 (1), 57–69.
Wu, J., Liu, M., Sun, H., Li, T., Wang, D., 2015. Equity-based timetable synchronization optimization in urban subway network. Transp. Res. C 51 (Feb.), 1–18.
Yang, X., Chen, A., Ning, B., Tang, T., 2016. A stochastic model for the integrated optimization on metro timetable and speed profile with uncertain train mass. Transp. Res. B 91, 424–445.
Yin, J., D'Ariano, A., Wang, Y., Xun, J., Su, S., Tang, T., 2019. Mixed-integer linear programming models for coordinated train timetabling with dynamic demand. In: IEEE Intelligent Transportation Systems Conference. pp. 863–868.
Yin, J., D'Ariano, A., Wang, Y., Yang, L., Tang, T., 2021. Timetable coordination in a rail transit network with time-dependent passenger demand. European J. Oper. Res. 295, 183–202.
Yin, J., Yang, L., D'Ariano, A., et al., 2022. Integrated backup rolling stock allocation and timetable rescheduling with uncertain time-variant passenger demand under disruptive events. INFORMS J. Comput. 34 (6), 3234–3258.
Yin, J., Yang, L., Tang, T., Gao, Z., Ran, B., 2017. Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. Transp. Res. B 97, 182–213.
Ying, C., Chow, A., Chin, K., 2020. An actor-critic deep reinforcement learning approach for metro train scheduling with rolling stock circulation under stochastic demand. Transp. Res. B 140, 210–235.