Article:

Wu, P.-Y., Guo, R.-Y., Lin, Z. et al. (2024) Improving passenger travel efficiency through a dynamic autonomous non-stop rail transit system. Transportation Research Part E: Logistics and Transportation Review, 182. 103414. ISSN: 1366-5545

https://doi.org/10.1016/j.tre.2024.103414

Improving passenger travel efficiency through a dynamic autonomous non-stop rail transit system

Pei-Yang Wu [a,b,c], Ren-Yong Guo [a,b,*], Zhiyuan Lin [c], Ronghui Liu [c], Pu Xu [a,b]

[a] *School of Economics and Management, Beihang University, Beijing 100191, China*

[b] *Key Laboratory of Complex System Analysis and Management Decision, Ministry of Education, Beijing 100191, China*

[c] *Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, United Kingdom*

Abstract: This study proposes a dynamic autonomous non-stop rail transit (DANRT) system to improve the travel efficiency of passengers in urban rail transit (URT) systems and solves a pertinent carriage scheduling problem derived from the DANRT system using a mathematical programming model. In the DANRT system, passengers traveling to the same destination are allocated to the same carriage(s), and each carriage can be attached to and detached from trains using the modular autonomous vehicle (MAV) technology effortlessly, which enables all trains to run non-stop throughout the focused operation period. We offer a cost-effective design for the DANRT system. To ensure safe and efficient operations, a mathematical model is proposed for the carriage scheduling problem in the DANRT system, where the number and destinations of carriages required by each station are determined. A linearization and segmentation method for the model is proposed. To examine the effectiveness of the DANRT system, we compare the travel efficiency of passengers in the DANRT system with that in the traditional system by using the origin-destination distribution of passengers on the Batong Line of Beijing Subway. The results demonstrate that passengers in DANRT can save about 2.9% to 8.6% of travel time compared with the traditional system. Finally, we conclude several observations and operational characteristics of the DANRT system by numerical experiments.

Keywords: Rail transit passengers; modification of transportation facility; modular autonomous vehicle; passenger travel efficiency; carriage scheduling optimization

---

[*] Corresponding author.

E-mail addresses: wupeiyang@buaa.edu.cn (P.-Y. Wu), guorenyong@buaa.edu.cn (R.-Y. Guo), Z.Lin@leeds.ac.uk (Z. Lin), R.Liu@its.leeds.ac.uk (R. Liu), xupu@buaa.edu.cn (P. Xu).

Improving passenger travel efficiency through a dynamic autonomous non-stop rail transit system

Abstract: This study proposes a dynamic autonomous non-stop rail transit (DANRT) system to improve the travel efficiency of passengers in urban rail transit (URT) systems and solves a pertinent carriage scheduling problem derived from the DANRT system using a mathematical programming model. In the DANRT system, passengers traveling to the same destination are allocated to the same carriage(s), and each carriage can be attached to and detached from trains using the modular autonomous vehicle (MAV) technology effortlessly, which enables all trains to run non-stop throughout the focused operation period. We offer a cost-effective design for the DANRT system. To ensure safe and efficient operations, a mathematical model is proposed for the carriage scheduling problem in the DANRT system, where the number and destinations of carriages required by each station are determined. A linearization and segmentation method for the model is proposed. To examine the effectiveness of the DANRT system, we compare the travel efficiency of passengers in the DANRT system with that in the traditional system by using the origin-destination distribution of passengers on the Batong Line of Beijing Subway. The results demonstrate that passengers in DANRT can save about 2.9% to 8.6% of travel time compared with the traditional system. Finally, we conclude several observations and operational characteristics of the DANRT system by numerical experiments.

Keywords: Rail transit passengers; modification of transportation facility; modular autonomous vehicle; passenger travel efficiency; carriage scheduling optimization

1. Introduction

As the urban population increases, the urban transportation system is under greater pressure. Urban rail transit (URT), as an important part of the urban transportation system, plays an increasingly significant role in meeting the travel demands of citizens and alleviating traffic congestion (Yuan et al., 2020). However, urban rail transit systems can no longer meet all the travel demands of citizens (Li et al., 2017). In urban rail transit stations, passengers will gather during rush hours. If the travel efficiency of passengers is low, congestion will appear in some stations, such as the Xierqi Station in Beijing, as shown in Fig. 1, and squeeze and stampede accidents are more likely to occur (Xu et al., 2016).

Fig. 1. A snapshot of passengers boarding at the Xierqi Station in Beijing. Some passengers are stuck at the door and a staff member marked in the red circle is pushing them into the train due to overcrowding in the carriage.

To alleviate congestion during rush hours, traffic demand management (TDM) strategies need to be adopted. TDM mainly includes pricing or fare-reward schemes (Kamel et al., 2020; Yang and Tang, 2018) and passenger flow control strategies (Yuan et al., 2020). The TDM strategies aim to shift some demands of passengers in busy stations in peak hours to off-peak hours or to delay passengers entering stations. Pricing or fare-reward schemes are difficult to implement due to the involvement of monetary compensation during their implementation (Shi et al., 2019). Thus, passenger flow control is a more popular method and is widely used in practice. Scholars always focus on the control of the inflow volume of passengers during a certain period at each station. For example, Yuan et al. (2020) aimed at minimizing the waiting time of passengers by proposing a mathematical programming model. Jiang et al. (2018) proposed a reinforcement learning-based method to minimize the penalty value of passengers being stranded. Two passenger flow control strategies were proposed in the above two studies (Jiang et al., 2018; Yuan et al., 2020). The congestion of passengers can be prevented by changing the demands of some passengers. However, the world urban population is projected to increase by more than 2 billion people between 2023 and 2050 (United Nations, 2019). With the growth of the urban population, changing the demands of passengers faces greater challenges, such as greater costs and fairness issues. In addition, if the demands of passengers from morning to night are always large in a URT system, the capacity of the URT system is not enough all the time even though the passenger demands are controlled. As a result, congestion will still appear in that case. Therefore, the performance of URT systems needs to be improved, i.e., traffic supply management (TSM) schemes are needed.

TSM mainly refers to the improvement of the performance of URT systems, such as the increase in the train capacity and the network expansion of URT systems (Liu et al., 2019). However, due to restrictions on land use and the cost of building extra urban rail lines, scholars pay more attention to optimizing train operations under an existing transportation system, such as optimizing the train timetables (Zhang et al., 2018) and the train stopping schemes (Jamili and Pourseyed Aghaee, 2015). In these studies of optimizing train operations, mathematical programming models are generally employed and the optimization objective can be to minimize the waiting time of passengers (Barrena et al., 2014; Shafahi and Khani, 2010), the energy consumption of train operations (Yin et al., 2016), the connection time between first trains (Guo et al., 2016). These optimization objectives can be greatly improved and the travel efficiency of passengers can be improved to some extent through optimizing train operations. In addition, several coordinated optimization schemes of passenger inflow control with train operation, such as train skip-stopping (Shi et al., 2023; Yuan et al., 2023) and train regulation (Li et al., 2017; Yuan et al., 2022), are also proposed. However, the travel efficiency of passengers cannot be further improved due to operational limitations, such as headway and train loading capacity (Shi et al., 2019).

In addition to the aforementioned traditional methods, scholars also propose several other methods to improve the operation of URT systems through retrofitting existing trains. For example, Daganzo (2022) introduced a kind of train that is substantially longer than the platforms to the urban railway system. Shi et al. (2022) explored a train operation method that allows carriages at different stations can be reserved via installing some isolation doors for a more reasonable distribution of train capacity to each station. These methods provide some new perspectives to improve the travel efficiency of passengers. Moreover, emerging new technologies, such as the modular autonomous vehicle (MAV) technology, make it possible to further improve the travel efficiency of passengers. Nold and Corman (2021) identified different generations of train coupling: Manual or automatic coupling/decoupling at stations can be identified as the first two generations. Virtual coupling (Felez et al., 2019) is recognized as the third generation of train coupling. The fourth generation of coupling refers to rail units that can mechanically attach and detach at operational speed automatically. MAV can be identified as the fourth generation of coupling (Nold and Corman, 2021).

Fig. 2 shows a class of MAVs. The MAV technology allows multiple identical pods (modular units) to be attached at speed on road as a modular vehicle and these pods can also be detached from others at full speed. Modular autonomous vehicles are being tested in Dubai[1] and Singapore[2]. Although the MAV is under development, it can be widely applied in the future, just like autonomous vehicles (Sala and Soriguera, 2021). Some scholars considered several optimization problems when the MAV technology is used (Chen and Li, 2021; Chen et al., 2022; Chen et al., 2020; Dai et al., 2020; Guo et al., 2017; Liu et al., 2023; Shi and Li, 2021; Tian et al., 2022; Zhang et al., 2020) as listed in Table 1. For instance, Chen and Li (2021) proposed a method to control the vehicle dispatch headway and vehicle capacity to solve the demand-supply asymmetry. Tian et al. (2022) developed an optimization model to determine the locations of stations enabling stationwise assembling/dissembling of modular units. Similarly, the focus of this paper is not the train control technologies and their feasibility. Rather, we assume that in the future, such technologies will be mature and we would like to investigate the optimization of the operations under such an assumption. It is meant for future scenarios. Different from Chen and Li (2021) and Tian et al. (2022), we focus on a new issue, i.e., the severe supply shortage during peak hours in URT systems, and will propose a novel dynamic autonomous non-stop rail transit (DANRT) system by using the MAV technology, more specifically the technology of automatic coupling and decoupling of carriages, and offer a cost-effective design for the system to improve the travel efficiency of passengers.

Table 1 Summary of studies on transportation systems when the MAV technology is used.

Publication	Background/ Issue addressed	System type	Passengers transfer between carriages	Direct destination for passengers
Guo et al. (2017)	Random demand	Many-to-one transit	–	–
Chen et al. (2020)	Demand–supply asymmetry	Transit shuttle	–	–
Dai et al. (2020)	Demand–supply asymmetry	Bus corridor	√	×
Zhang et al.	Demand–supply	Bus network	√	√

[1] https://www.reuters.com/article/us-emirates-transportation-autonomous/dubai-tests-autonomous-pods-in-drive-for-smart-city-idUSKCN1GD5G6
[2] https://www.next-future-mobility.com/

(2020)	asymmetry and first- and last-mile problem			
Chen and Li (2021)	Demand–supply asymmetry	Transit corridor	√	×
Shi and Li (2021)	Demand–supply asymmetry and oversaturated queuing phenomenon	Transit corridor	√	×
Chen et al. (2022)	Demand–supply asymmetry	Transit corridor	√	×
Tian et al. (2022)	Stations location problem	Transit corridor	√	×
Liu et al. (2023)	Demand–supply asymmetry	Bus loop line	√	×
This paper	Severe supply shortage	Transit corridor	×	√

Note: '–', '√', and '×' represent that a passenger movement rule is not involved, is permitted, and is not permitted, respectively.

(a) (b)

Fig. 2. A class of modular autonomous vehicles. (a) One pod. (b) A modular vehicle with two connected pods. Source: https://www.next-future-mobility.com/.

In a DANRT system, each carriage can be attached to and detached from a train easily as shown in Fig. 3 (corresponding to the word 'autonomous' in the name 'DANRT'). In each station, there are some vacant carriages for passengers to board. Passengers in the same carriage travel to the same destination. Carriages occupied by passengers will be attached when a train passes through the station and detached after arriving at the corresponding destination. The train does not stop all the way (corresponding to the word ‘non-stop’ in the name ‘DANRT’).

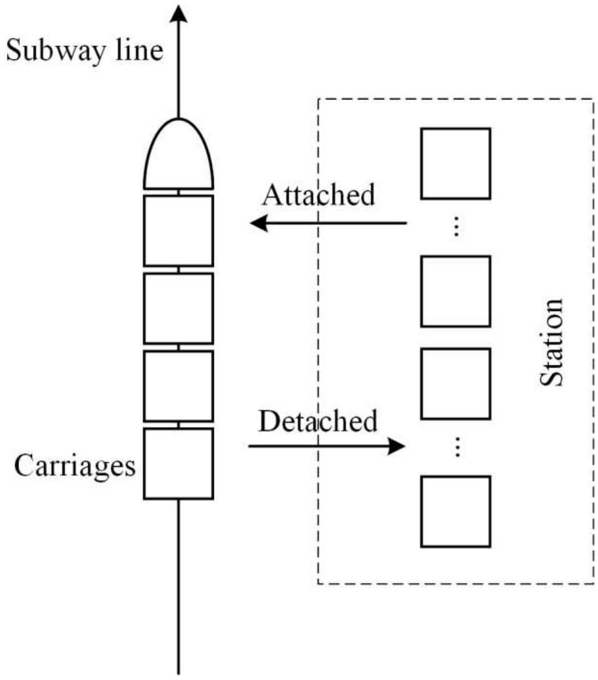


Fig. 3. A schematic diagram of the DANRT system.

In a DANRT system, the length of each carriage is reduced compared to traditional URT systems. Consequently, the DANRT system can accommodate a greater number of carriages per train. A train slows down as it runs close to a station, maintains a lower speed until all the carriages are detached and attached, and then speeds up again without stopping. Near each station, there is an extra train track line to ensure that the train can keep running all the way. The carriages needed in each station are guaranteed by scheduling (corresponding to the word ‘dynamic’ in the name ‘DANRT’). There is a screen on each carriage to display its destination. Upon entering a station, passengers will receive location information for a carriage that corresponds to their destination either through their mobile phones or via a large screen display at the station. They can then proceed to board the designated carriage. When a train passes through the station, the carriage is attached to the train. When the train passes through the destination, the carriage is detached and parked at the platform. The passenger gets off the carriage and reaches his/her destination. The train does not stop during the whole process. In this way, the travel efficiency of passengers can be improved due to the non-stop design.

We offer a possible design of the system so as to simultaneously keep trains running all the way, ensure that passengers can get on and off carriages safely, and save the construction cost of stations as much as possible. The second problem we need to address is the carriage scheduling problem. The operation of a DANRT system depends on the scheduling of carriages. A better carriage scheduling scheme leads to a higher travel efficiency of passengers under the premise of safety. Therefore, we need to obtain a carriage scheduling scheme, including the number and destinations of carriages required by each station, to ensure safe and efficient operations of the system. Considering the physical limitations of the stations and carriages, we formulate the carriage scheduling problem in the DANRT system as a nonlinear integer programming model. A linearization and segmentation method is proposed to facilitate model solving. The third problem we want to answer is what the characteristics of the DANRT system are compared with a traditional URT system, such as the travel efficiency of passengers and the necessity of the existence of storage lines. We propose a mathematical model of the traditional URT system for comparison. We use the origin-destination (OD) distribution of passengers on the Batong Line of Beijing Subway as model input and conclude the characteristics of the DANRT system by adjusting the key parameters in the model.

The DANRT system is different from the Personal Rapid Transit (PRT) (Shafahi and Khani, 2010; Schweizer et al., 2011). In a PRT system, small vehicles are always used to cater to an individual or a small group. As a result, the capacity of each vehicle is not effectively utilized and the operation cost seems to be large (Ceder, 2021). However, the goal of a DANRT system is not to ensure a better user experience as a PRT system does, but to improve travel efficiency. A DANRT system is usually applied to a URT system where there is a severe shortage of supply. There are a large number of passengers in URT stations, and thus the capacity of each carriage can be utilized more effectively. The DANRT system is also different from the corridor systems using MAV technology (Chen and Li, 2021; Chen et al., 2022; Tian et al., 2022) due to the following three reasons. First, this paper aims to address the problem of the severe supply shortage in the URT system and improve the travel efficiency of passengers. However, Chen and Li (2021) and Chen et al. (2022) aimed at solving the demand-supply asymmetry and Tian et al. (2022) aimed to determine the locations of stations and proposed operation strategies. Therefore, we consider the scenarios in which passenger demands are extremely large and tend to explore the potential of the proposed DANRT system in improving the travel efficiency of passengers. Second, we offer a novel design of the DANRT system, which is different from the corridor systems (Chen and Li, 2021; Chen et al., 2022; Tian et al., 2022). (1) It is difficult for passengers in carriages to move to the adjacent carriages during morning or evening rush hours because there may be too many passengers in a carriage as shown in Fig. 1. Thus, passengers in the same carriage travel to the same destination and are prohibited from moving to the adjacent carriages in the DANRT system. (2) Trains keep running all the way instead of stopping at each station and thus passengers can directly reach their destinations as listed in Table 1. Third, we address the carriage scheduling problem in the DANRT system to optimize the operation of the system.

Our contributions are as follows. (1) We focus on the issue of the severe supply shortage during peak hours in URT systems and propose a DANRT system to improve the travel efficiency of passengers in URT systems. Passengers in the DANRT system are prohibited from moving to the adjacent carriages and can directly reach their destinations. A cost-effective design for the DANRT system is also proposed. (2) The problem of carriage scheduling during train operations is formulated as a nonlinear integer programming model. The model can be easily extended to other DANRT systems. A linearization and segmentation method for the model is proposed to facilitate model solving. (3) We demonstrate the effectiveness of the DANRT system by using the origin-destination distribution of passengers on the Batong Line of Beijing Subway in October 2017. Meanwhile, observations and operational characteristics of the DANRT system are found through numerical experiments.

The remaining of this paper is arranged as follows. In Section 2, we present a possible design of the DANRT system and introduce the train operation process of the DANRT system. In Section 3, a mathematical model is proposed to address the carriage scheduling problem in the DANRT system. A mathematical model of a traditional URT system is also presented for comparison. Meanwhile, a linearization and segmentation method for the model is proposed to facilitate model solving. Section 4 offers a set of numerical results and illustrates the effectiveness of the DANRT system. Some insightful characteristics of the DANRT system are provided. Section 5 concludes this paper.

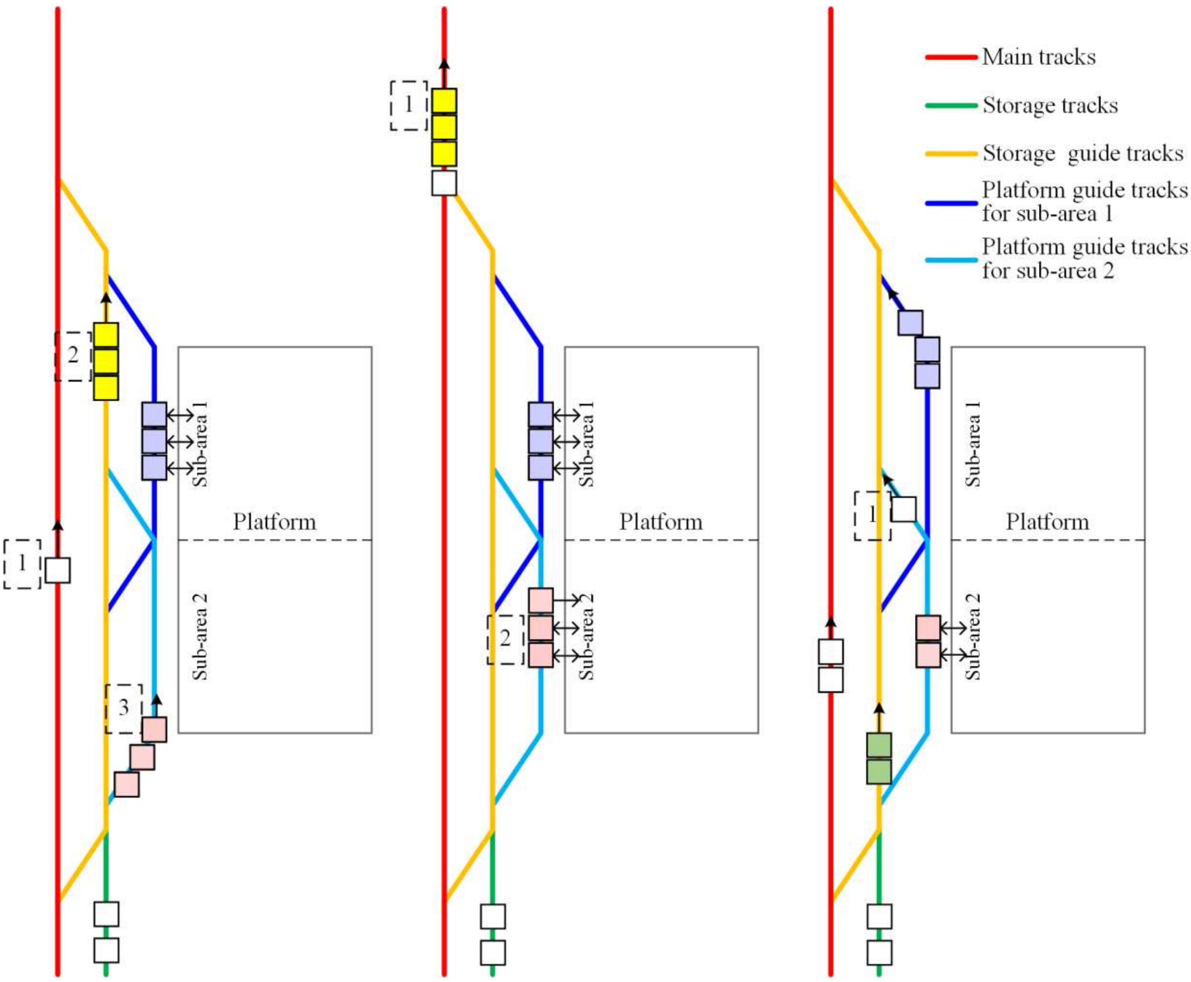
2. DANRT system

In this section, we offer a possible design of the DANRT system. Certain operational details of trains are presented in Appendix A, demonstrating the feasibility of implementing such a DANRT system. In addition, we illustrate the advantages and disadvantages of the DANRT system based on the design, from both technical and economic perspectives. We only consider one direction of a bidirectional line. The design method for the other direction of the bidirectional line is the same.

2.1. A possible design of the DANRT system

If we assume that the time required for passenger boarding and alighting (i.e., boarding time plus alighting time) is t_s, the headway between successive trains in the traditional system has to be at least t_s. We assume that the headway between successive trains in the traditional system takes the minimum value t_s for the highest travel efficiency, and it takes $t_s/2$ for passengers to get off a carriage and $t_s/2$ for passengers to get on a carriage. To increase the capacity of the system, we propose a shorter headway $t_s/2$, which requires modifications on the station layout. Therefore, we present a design of the DANRT system, which is called a **double-area system**. The minimum headway between successive trains in the double-area system is at least $t_s/2$. However, in a DANRT system, trains on main tracks will slow down when passing through a station. In order to avoid rear-end collisions, the headway may be larger than $t_s/2$. In addition, the minimum headway should also include the time taken to pull in and pull out the carriages to ensure that passengers have enough time to get on and off a carriage. As a result, the minimum headway should be larger than $t_s/2$ in the double-area system. These times will be considered when setting the minimum headway in Section 4. For more details on minimum headway, please refer to Appendix A.

As shown in Fig. 4, the red lines, green lines, and orange lines represent the main tracks, storage tracks, and storage guide tracks, respectively. The platform of the station is dynamically divided into two sub-areas. Two successive trains will be parked at different sub-areas. The platform guide tracks for sub-area 1 are denoted by dark blue lines and the platform guide tracks for sub-area 2 are denoted by sky blue lines. Carriages are represented by squares. It is worth mentioning that the boundary between sub-areas 1 and 2 is not necessarily the center line of the platform as shown in Fig. 4. The size of each sub-area can change when a train passes through the station. There may be more guide tracks between the main tracks and the platform accordingly.



(a) (b) (c)

Fig. 4. The operation process of carriages in the double-area system when (a) a train is passing through a station, (b) the train is about to leave the station, and (c) the next train is passing through the station.

When a train passes through the station (see position 1 in Fig. 4(a)), carriages arriving at their destination are detached and parked at sub-area 2 (see position 3 in Fig. 4(a) and position 2 in Fig. 4(b)). The carriages parked at sub-area 2, which are represented by yellow squares, are attached to the train along the guide tracks (see position 2 in Fig. 4(a) and position 1 in Fig. 4(b)). The carriages parked at sub-area 1, which are represented by light blue squares, have been parked for $t_s/2$, i.e., the headway between successive trains in the double-area system. When the next train passes through the station, the carriages parked at sub-area 1 will have been parked for t_s which is enough for passengers to complete their boarding and alighting process. Subsequently, the carriages parked at sub-area 1 will be attached to the next train and the carriages detached from the next train will be parked at sub-area 1. Appendix B provides a more detailed operation process of carriages passing through a station in the double-area system.

If the number of carriages arriving at their destination is more than the number of carriages needed in the station, a wait for all the passengers in a certain number of carriages to get off is needed. These carriages become vacant and can be either attached to the next train or parked on storage tracks. For example, we assume that three carriages are detached from the train. However, the station only needs two. A wait for the passengers in a carriage to get off is needed. The vacant carriage can be then attached to the next train as shown in Fig. 4(c) or parked on storage tracks. On the other hand, if the number of carriages arriving at their destination is fewer than the number of carriages needed in the station, a certain number of vacant carriages can be either detached from the train or transferred from storage tracks in the station.

We will illustrate that the storage lines in the DANRT system are unnecessary when we use the OD distribution of passengers on the Batong Line of Beijing Subway as model input in Section 4.2.

2.2. Discussions and implications of DANRT

Compared with a traditional URT system, a DANRT system has two advantages in improving travel efficiency. First, the travel time of passengers on trains will be reduced in a DANRT system. We assume that the time required for boarding and alighting is t_s. In traditional URT, a train has to be parked at each station for at least t_s. In contrast, trains keep running all the way in DANRT. Consequently, passengers traveling on a DANRT train can save a considerable amount of time as they pass through each station. For more in-depth information, please refer to Appendix A. Second, under DANRT, there is a possibility to save the total waiting time of passengers on platforms. In traditional URT, passengers get on the train in a first-come-first-served manner, which fails to minimize the total waiting time of passengers on platforms. The reason is that the total waiting time of passengers on platforms is related to the sequence of passengers entering the station. For example, passengers A and B enter the starting station almost at the same time. We assume that only one passenger can get on the train. The destination of passenger A is the next station and the destination of passenger B is the terminal station. The traditional URT will allocate the one arriving slightly earlier to the train. If passenger A enters the station first, he/she will get off the train at the next station and one more passenger in the next station will get on the train. As a result, the total waiting time of passengers will be shorter in this case. First-come-first-served manner does not guarantee a minimum passenger waiting time. In DANRT, however, a passenger has to get on a carriage corresponding to his/her destination. If the carriages corresponding to his/her destination are full, he/she has to wait for the next train, even if the rest of the carriages to the other destinations are vacant. Ostensibly, this may lead to underutilization of carriage capacity. However, we point out that this design actually makes it possible to control the number of passengers by taking into account their varied destinations, so as to reduce the total passenger waiting time. For example, we assume that a carriage can load 10 passengers. 10 passengers whose destination is station 3 (terminal station) are waiting at station 1 (starting station), while 10 passengers whose destination is station 2 are also waiting at station 1. In this case, if we give priority to providing a carriage for the 10 passengers whose destination is station 2 instead of those passengers whose destination is station 3, the carriage will be detached from the train when the train passes through station 2 and one more carriage at station 2 can be attached to the train. As a result, more passengers at station 2 who intend to go to station 3 can get on the train and the waiting time of passengers at station 2 will be shorter. Meanwhile, the waiting times of passengers at station 1 in the two scenarios (prioritizing passengers going to stations 2 and 3 respectively) are the same. Thus, the total passenger waiting time will be shorter if we prioritize passengers going to station 2.

This design may bring about some fairness issues. For example, passengers who intend to go to nearer stations or more popular stations are more likely to get on a train. However, some measures, e.g., providing a class of flexible carriages (see the last paragraph in this section) and setting appropriate objective functions (see Section 3), can be adopted to solve the fairness issues to some extent. Furthermore, this design may initially raise concerns about passengers needing additional time to locate a carriage that corresponds to their destination. However, each passenger will receive precise location information for their desired carriage via their mobile phone or the large screen at the station before reaching the platform. Consequently, passengers can simply proceed directly to the designated location without any extra time requirement. Moreover, passengers who prefer not to receive this information have the option to choose a flexible carriage that stops at every station. Essentially, in a DANRT system, passengers have an added level of choice.

Another management aspect to consider is that passengers in a DANRT system may need to walk longer distances to access a carriage that matches their destination. Nevertheless, it is worth noting that this issue is not unique to DANRT systems and exists in traditional systems as well. During peak passenger flow, many passengers tend to gather near entrances, resulting in limited capacity in carriages near these entry points shortly after a train arrives. Consequently, passengers who have just entered the platform often need to walk considerable distances to find a carriage with available space and board the train. However, in a DANRT system, it becomes possible to alleviate congestion near entrances. For instance, some carriages corresponding to popular stations can be strategically positioned further away from the entrances, thereby distributing passenger load more evenly. After all, when there is congestion near entrances, it is more important to alleviate congestion (e.g., distribute passenger load more evenly) than to allow some passengers to walk less distance.

Furthermore, within a DANRT system, the carriage capacity could remain underutilized, leading to vacant seats in certain carriages during peak hours. Passengers might initially feel confused upon encountering these unoccupied seats when facing overcrowding. This scenario has the potential to affect service satisfaction. However, we will show that passengers as a whole can benefit from this system design in Section 4. Moreover, this problem can also be alleviated or even solved from at least two aspects. First, a DANRT system is more suitable for managing large demand, especially for unbalanced demand (we will show this in Section 4). In this case, the capacity of each carriage can be more effectively utilized. Therefore, a DANRT system should be applied to a URT system with large and unbalanced passenger demand to minimize the occurrence of vacant seats in carriages. When passenger demand is small or balanced, a DANRT system can be transformed into the traditional one to avoid the occurrence of this problem. Second, after solving the carriage scheduling problem, we can check the obtained carriage scheduling scheme. If there are carriages with (a large number of) vacant seats inside, these carriages can be set as flexible carriages that are available to everyone to improve service satisfaction.

In terms of technical feasibility, the MAV technology can be implemented here to ensure safe and seamless coupling/decoupling operations on carriages. Each carriage has its own power supply and can be regarded as a modular autonomous vehicle. Moreover, in rail transit system, a possible method to realize coupling and decoupling of carriages traveling at operational speed is also proposed (Nold, 2019; Nold and Corman, 2021): Carriages can run with small distance separation or be connected as a platoon (Bhoopalam et al., 2018; Boysen et al., 2018; Schwerdfeger et al., 2021). When carriages are particularly close, a railway coupling support device (RCSD) and coordinated communication would be required. Technically, all those technologies have been shown separately feasible (Nold and Corman, 2021). In addition, the operating time of a conventional electric switch machine can be no more than 0.6 s (Qian et al., 2019). Therefore, when a train passes a station at a lower speed (e.g., 5 m/s), each carriage, including the carriages in the middle of the train, can be safely decoupled by keeping a certain distance (e.g., 3 m) from the adjacent carriage that does not need to be decoupled in advance. Vehicle-based switching (Farhat et al., 2018; Goodall, 2010) can be implemented here to further improve ride quality. In other words, although the MAV and RCSD are under development, it is feasible to implement these technologies and apply them in rail transit system in the near future. These technologies are summarized in Table 2.

Table 2 Futuristic technologies required.

Technologies	Descriptions
MAV or RCSD	Ensure safe and seamless coupling/decoupling operations on carriages at operational speed. To implement these two technologies, coordinated communication would be required.
Unmanned driving	There are a large number of carriages in a DANRT system. Therefore, it is unlikely that there will be a driver in each carriage. Unmanned driving would be required in a DANRT system.
Vehicle-based switching	Switch routes from on-board the vehicle. The wheels can be individually controlled with torque-controlled motors relying on sensors. This technology ensures that carriages can quickly switch to another track line. Although traditional conventional electric switch machines can also be used here, this technology can further improve ride quality.

In terms of economic feasibility, only an extra train track line and some guide tracks are required near each station in a DANRT system. Thus, the land use, construction cost of stations, and reconstruction of carriages are acceptable compared to the benefits brought by the system. Note that a DANRT system is suitable for application in urban rail lines/systems where there is a severe shortage of supply. It is worthwhile or even necessary to make a one-time investment to improve the travel efficiency of almost all passengers and reduce the risk of stampedes in those lines/systems in the coming decades. In those lines/systems, the passenger flow may not always be huge. Therefore, different operational strategies can be adopted. When the passenger flow is extremely low, due to the small-carriage design, customized services can be provided in a DANRT system, which can meet the demand of passengers for customized services. When the passenger flow is moderate, a DANRT system can be transformed into the traditional one, i.e., a certain number of carriages are connected as a train that stops at each station, to save energy consumption due to capacity flexibility of trains. In addition, when the passenger flow is low or moderate, a DANRT system can be combined with logistics systems (Behiri et al., 2018; Visser, 2018; Zhao et al., 2021) to improve economic benefits, i.e., vacant carriages can be used for freight. Specifically, a proportion of carriages are used to accommodate passengers and the remaining carriages are used to transport freight. Note that the economic issues are not the focus of this paper. It can be assumed that if MAV and RCSD are mature technologies in the future, investments on new stations accommodating DANRT will be well accepted.

It is worth mentioning that we do not force passengers in carriages to move to adjacent carriages to alight as in Tian et al. (2022). The reason is that a DANRT system is always applied to URT lines with huge passenger flow. In this case, it may be difficult for passengers in carriages to move. Therefore, when a carriage arrives at the corresponding destination in a DANRT system, the carriage is detached directly. In addition, in reality, there may be a small number of passengers who enter a station without determining their destinations. For those passengers, a DANRT system can provide a class of flexible carriages that stop at each station. Specifically, a certain part of parking space in each station is reserved for the flexible carriages that stop at each station. Meanwhile, a certain part of the loading capacity is also reserved by trains. As a result, the flexible carriages in the first station can be coupled to a train and decoupled at the second station. These carriages are then coupled to the next (or third) train and decoupled at the third station. In this way, passengers in those carriages can get off at any station. The existence of flexible carriages can solve the fairness issues to some extent: Those passengers who are not prioritized can choose the flexible carriages.

3. Models

This study considers one direction of a bidirectional line and an ordered set of trains running in the DANRT system (i.e., the double-area system). We aim to obtain a carriage scheduling scheme for the DANRT system to maximize the travel efficiency of passengers in a certain time period. A carriage scheduling scheme includes the number and destinations of carriages needed in each station when a train passes through the station. Additionally, to meet the demand of each station for carriages, we also need to obtain the number of vacant carriages attached to each train from an upstream station to a downstream station.

We use the total waiting time of passengers on platforms as the indicator to reflect the travel efficiency of passengers. One can also use the capacity of the DANRT system or the total time saved by passengers on trains as the indicator. However, using the last indicator may lead to some fairness issues. Specifically, passengers who intend to go to further stations are more likely to get on a train. The reason is that each passenger can save a certain amount of time when he/she passes through each station, as mentioned in Section 2.2, and thus passengers who intend to go to further stations will save more time. However, if we use the total waiting time of passengers on platforms as the indicator, we can give the waiting time of passengers who have waited longer a higher weight to make them get on the next train more likely. If the capacity of the DANRT system is used as the indicator, we can also give the number of passengers who have waited longer a higher weight to make them get on the next train more likely. Managers can solve the fairness issues to some extent by adjusting this weight. Note that passenger waiting cost is always far higher than the train operation cost (Chen and Li, 2021), and thus we do not consider the train operation cost here.

We propose a mathematical model to address the carriage scheduling problem in the double-area system. Additionally, to show the effectiveness of the DANRT system in improving the travel efficiency of passengers, we also need to obtain the total waiting time of passengers in a traditional URT system for comparison. However, in the traditional URT system, passengers get on the train in a first-come-first-served manner, and hence the total waiting time of passengers on platforms is related to the sequence of passengers entering the station. Different passenger arrival sequences have different effects on the total waiting time of passengers. There is an optimal arrival sequence of passengers in the traditional system which can minimize the total waiting time. To better demonstrate the effectiveness of the DANRT system, we should compare the total waiting time of passengers in the double-area system with the theoretical minimum total waiting time of passengers in the traditional system. Therefore, we propose a model so as to obtain the theoretical minimum total waiting time of passengers in the traditional system by solving the model. We set the OD demands of passengers in the traditional system to be the same as those in the DANRT system when solving the model. In this way, the effect of the sequence of passengers entering the station on the indicator is eliminated. We can simultaneously obtain the optimal passenger arrival sequence in the traditional system by solving the model. That is to say, the optimal passenger arrival sequence is not the input of the traditional system model, but a result from solving the model. If the passenger arrival sequence is determined, the total waiting time can be directly obtained without solving the model. It is worth mentioning that the total waiting time of passengers obtained by solving the model of the traditional URT system must be shorter than the actual total waiting time of passengers in the traditional URT system. If we apply the input data with any other passenger arrival sequence in the traditional system, the total waiting time of passengers will be longer than the theoretical minimum total waiting time.

Therefore, the effectiveness of the DANRT system can be demonstrated by comparing the total waiting time of passengers obtained by solving the model of the DANRT system with that obtained by solving the model of the traditional system.

In Section 3.1, we formulate the carriage scheduling problem in the double-area system as a mathematical model. In Section 3.2, we develop a mathematical model to eliminate the effect of passenger arrival sequence on the total waiting time of passengers in the traditional system. A linearization and segmentation method is proposed to facilitate model solving in Section 3.3. The indices, parameters, and variables used throughout this paper are shown in Table 3.

Table 3 Sets, subscripts, parameters, and decision variables.

Notations	Detailed definitions
Sets and subscripts	
i , j	Indices of stations in the system, $i, j = 1, 2, \cdots, m$.
S	The set of OD pairs, $S = \{(i, j) \mid i = 2, 3, \cdots, m, j \le i\}$.
$k \in K$	The index and set of trains, $K = \{1, 2, \cdots, k_{\max}\}$.
C	The set of numbers of carriages required by stations to accommodate passengers, $C = \{c_{ij}^k \mid (i, j) \in S, k \in K\}$.
V	The set of numbers of vacant carriages detached from by trains, $V = \{v_{ij}^k \mid (i, j) \in S, k \in K\}$.
R	The set of numbers of carriages on storage lines, $R = \{r_j^k \mid j = 2, 3, \cdots, m, k = 0, 1, \cdots, k_{\max}\}$.
$\tilde{C}$	The set of the number of passengers who can get on trains in traditional URT, $\tilde{C} = \{\tilde{c}_{ij}^k \mid (i, j) \in S, k \in K\}$.
Parameters	
t	The headway between successive trains (unit: s).
t_s	The time required for the boarding and alighting process of passengers (unit: s).
m	The number of stations in the URT line.
$k_{\max}$	The number of trains running in the system.
T_{tra}	The maximum number of carriages a train can load at the same time.

T_{pla}	The maximum number of carriages that can be parked at a platform at the same time.
n	The maximum number of carriages that can be parked on a storage line at the same time.
c_{capa}	The capacity of a carriage.
x_{ij}^{k}	The number of passengers who just arrive at the platform of station i and also intend to go to station j in the interval between the arrival of train $k-1$ and train k.
γ	The maximum number of carriages in the system during the study period.
T_{capa}	The capacity of a train in traditional URT.
$Redu$	Parameter to depict the increase in actual OD demands of passengers.
δ	The penalty coefficient for waiting time of passengers who cannot get on the current train.

Decision variables

h_{ij}^{k}	The number of passengers on the platform of station i and also intend to go to station j when train k arrives. These passengers not only include the ones who just arrive at the platform but also include the remaining passengers on the platform who have not got on the previous train.
c_{ij}^{k}	The number of carriages that is required by station i to accommodate passengers who intend to travel from station i to station j when train k arrives.
v_{ij}^{k}	The number of vacant carriages detached from train k from station i to station j.
r_{j}^{k}	The number of carriages on the storage line at station j after train k arrives for $k=1,2,\cdots,k_{\max}$.
r_{j}^{0}	The initial number of carriages on the storage line at station j.
$\tilde{c}_{ij}^{k}$	The number of passengers who can get on train k at station i and intend to go to station j in the traditional system.
T_{rem}^{ik}	The remaining capacity of train k when the train is about to arrive station i.

We consider a subway line with m stations, including a starting station (indexed by m), a terminal station (indexed by 1), and $k_{\max}$ trains running in the system. The headway between successive trains is set to t. We use the concept of equivalent time (Sun et al., 2014; Yuan et al., 2020) here: We define time interval k as the interval between the departure of train $k-1$ and train k as shown in Fig. 5. Time interval 1 (with the value t) means the time span before train 1 is due to depart. We investigate a time span of $k_{\max} \cdot t$ which is evenly divided into $k_{\max}$ time intervals. Note that time interval k in the traditional system is different from that in the double-area system even if the departure intervals in the two systems are the same due to the non-stop design of trains in the DANRT system as shown in Fig. 5. In Fig. 5, trains in the double-area system and the traditional system depart from the starting station (station 12) at the same time and the departure intervals in the two systems are equal. Train 1 in the double-area system arrives at station 10 at t_1 and leaves the station without stopping. However, train 1 in the traditional system arrives at station 10 at t_2 and leaves the station at t_3. In other words, the timetables in the two systems are different, and thus the time intervals in the two systems are different. In addition, one can see that each passenger at station 12 who intends to go to station 10 will save $t_2 - t_1$ if he/she chooses the double-area system instead of the traditional system.

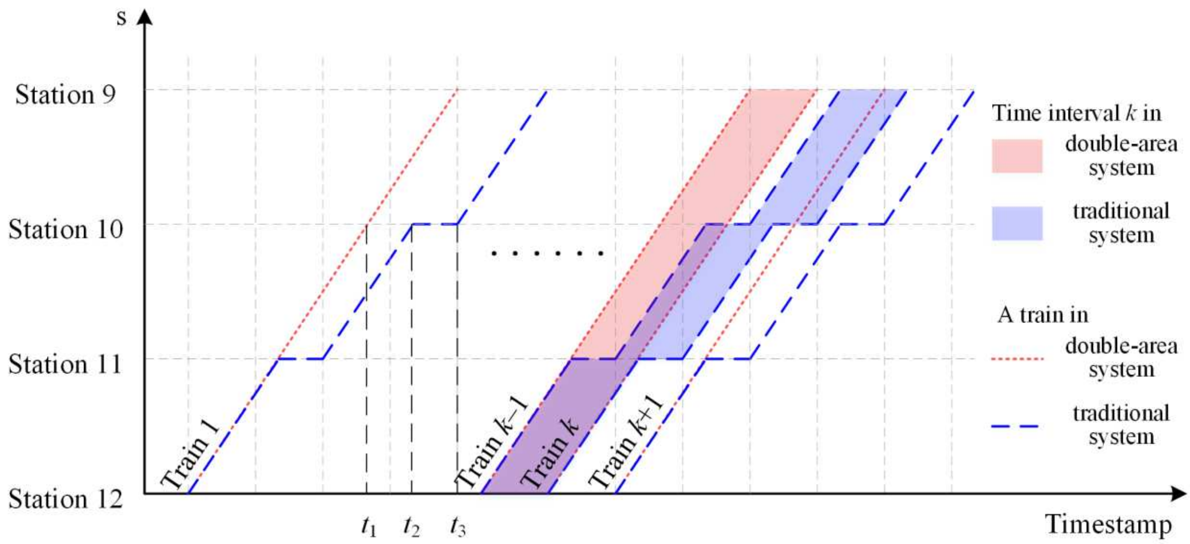


Fig. 5. A schematic diagram of timetables in the double-area system and the traditional system when the departure intervals in the two systems are the same.

3.1. Model of the double-area system

Some assumptions are adopted here. (1) Passenger demands represented by a set of OD matrices are known. In fact, we can obtain these OD matrices through a passenger demand forecast method (Yang et al., 2021). A travel reservation system (Han et al., 2020) can also be applied here to obtain these OD matrices. In other words, the proposed model is for planning-phase planning. If some passenger demands are not obtained in reality, some flexible carriages can be added. Note that we want to focus on the original DANRT system, and thus the passengers who enter a station without determining their destinations and the flexible carriages that stop at each station (see Section 2.2) are not considered in the model. If the flexible carriages need to be considered, it is sufficient to reduce T_{tra} and T_{pla} to model these carriages. For example, if the number of flexible carriages is set to 2, T_{tra} and T_{pla} in Sections 3.1 and 3.2 should be set to $T_{\text{tra}}-2$ and $T_{\text{pla}}-2$, respectively and T_{capa} in Section 3.3 should be set to $T_{\text{capa}}-T_{\text{capa}}\cdot 2/T_{\text{pla}}$. (2) The average waiting time of passengers arriving during the current time interval is measured by $t/2$, while the average waiting time of passengers failing to board the train during the previous time interval is measured by t. This assumption has been used widely in waiting time cost estimation in the case that service is regular (Ansari Esfeh et al., 2021; Dai et al., 2020; Tian et al., 2012). (3) Only the starting and terminal stations have garages that can store a large number of vacant carriages. There is a limit to the number of vacant carriages that can be stored on storage lines in the other stations.

Applying those notations summarized in Table 3, we formulate the carriage scheduling problem in the double-area system as the following optimization model:

$$\min_{(C,V,R)} \sum_{k\in K}\sum_{(i,j)\in S}\left(\delta\cdot\left(h_{ij}^{k}-x_{ij}^{k}\right)\cdot t+x_{ij}^{k}\cdot\frac{t}{2}\right), \tag{1}$$

$$s.t.\quad \sum_{i=\tilde{h}}^{m}\sum_{j=1}^{\tilde{h}-1}\left(c_{ij}^{k}+v_{ij}^{k}\right)\le T_{\text{tra}},\ \tilde{h}=2,3,\cdots,m,\ k\in K, \tag{2}$$

$$\sum_{i=j+1}^{m} c_{ij}^{k}\le T_{\text{pla}}-1, j=1,2,\cdots,m-1,\ k\in K, \tag{3}$$

$$\sum_{i=1}^{j-1} c_{ji}^{k}\le T_{\text{pla}}-1,\ j=2,3,\cdots,m,\ k\in K, \tag{4}$$

$$\sum_{i=j+1}^{m} c_{ij}^{k-1}+\sum_{i=1}^{j-1} c_{ji}^{k}\le T_{\text{pla}}-1,\ j=2,3,\cdots,m-1,\ k=2,3,\cdots,k_{\max}, \tag{5}$$

$$r_j^k = \begin{cases} r_j^{k-1} + \sum_{i=j+1}^{m} v_{ij}^k - \sum_{i=1}^{j-1} v_{ji}^k - \max\left\{\sum_{i=1}^{j-1} c_{ji}^{k+2} - \sum_{i=j+1}^{m} c_{ij}^k, 0\right\} \\ \qquad + \max\left\{\sum_{i=j+1}^{m} c_{ij}^{k-2} - \sum_{i=1}^{j-1} c_{ji}^k, 0\right\}, & k = 3,4,\cdots,k_{\max}-2,\ j = 2,3,\cdots,m-1, \\ r_j^{k-1} + \sum_{i=j+1}^{m} v_{ij}^k - \sum_{i=1}^{j-1} v_{ji}^k - \max\left\{\sum_{i=1}^{j-1} c_{ji}^{k+2} - \sum_{i=j+1}^{m} c_{ij}^k, 0\right\}, & k = 1,2,\ j = 2,3,\cdots,m-1, \\ r_j^{k-1} - \left(\sum_{i=1}^{j-1} c_{ji}^{k+2} + \sum_{i=1}^{j-1} v_{ji}^k\right), & k = 1,2,\cdots,k_{\max}-2,\ j = m, \\ r_j^{k-1} + \sum_{i=j+1}^{m} v_{ij}^k - \sum_{i=1}^{j-1} v_{ji}^k + \max\left\{\sum_{i=j+1}^{m} c_{ij}^{k-2} - \sum_{i=1}^{j-1} c_{ji}^k, 0\right\}, & k = k_{\max}-1, k_{\max},\ j = 2,3,\cdots,m-1, \\ r_j^{k-1} - \sum_{i=1}^{j-1} v_{ji}^k, & k = k_{\max}-1, k_{\max},\ j = m, \end{cases} \tag{6}$$

$$r_j^k \le n,\ k = 0,1,\cdots,k_{\max}-2,\ j = 2,3,\cdots,m-1, \tag{7}$$

$$r_j^{k_{\max}-1} \le n + T_{\text{pla}} - 1 - \sum_{i=1}^{j-1} c_{ji}^{k_{\max}},\ j = 2,3,\cdots,m-1, \tag{8}$$

$$r_j^{k_{\max}} \le n + T_{\text{pla}} - 1,\ j = 2,3,\cdots,m-1, \tag{9}$$

$$h_{ij}^1 = x_{ij}^1,\ (i,j) \in S, \tag{10}$$

$$h_{ij}^k = h_{ij}^{k-1} + x_{ij}^k - \min\left\{h_{ij}^{k-1}, c_{ij}^{k-1} \cdot c_{\text{capa}}\right\},\ (i,j) \in S,\ k = 2,3,\cdots,k_{\max}, \tag{11}$$

$$\sum_{j=2}^{m} r_j^0 + \sum_{k=1}^{2} \sum_{(i,j)\in S} \left(c_{ij}^k + v_{ij}^k\right) \le \gamma, \tag{12}$$

$$c_{ij}^k, v_{ij}^k \in N,\ (i,j) \in S,\ k \in K, \tag{13}$$

$$r_j^k \in N,\ k = 0,1,\cdots,k_{\max},\ j = 1,2,\cdots,m. \tag{14}$$

The optimization model minimizes the total waiting time of passengers on platforms within a time span of $k_{\max} \cdot t$. The total waiting time consists of the waiting time of passengers arriving during the current time interval and the waiting time of passengers failing to board the train during the previous time interval at each station for each train. x_{ij}^k denotes the number of passengers who just arrive at the platform of station i and also intend to go to station j in the interval between the arrival of train $k-1$ and train k. $h_{ij}^k - x_{ij}^k$ stands for the number of passengers who fail to board the train during the previous time interval at station i and also intend to go to station j when train k passes through the station. The average waiting time of these passengers is measured by t. δ (≥ 1) is a penalty coefficient for the waiting time of passengers who cannot get on the current train.

Constraint (2) guarantees that the number of carriages attached to train k after the train passes through station $\tilde{h}$, including both occupied and vacant carriages, does not exceed the maximum number of carriages that the train can load. $\sum_{j=1}^{\tilde{h}-1}\left(c_{ij}^{k}+v_{ij}^{k}\right)$ means the number of carriages that are attached to train k from station i and will go to the downstream stations of station $\tilde{h}$ (i.e., stations $1,2,\cdots,\tilde{h}-1$). The number of carriages that are coupled to train k from the upstream stations of station $\tilde{h}$ (including station $\tilde{h}$) and will go to the downstream stations of station $\tilde{h}$ is the number of carriages attached to train k after the train passes through station $\tilde{h}$. For example, we assume that there is a line with 4 stations and one train running in the system. Station 4 is the starting station and station 1 is the terminal station. It is assumed that $c_{43}^{1}+v_{43}^{1}=2$, $c_{42}^{1}+v_{42}^{1}=3$, $c_{41}^{1}+v_{41}^{1}=1$, $c_{32}^{1}+v_{32}^{1}=4$, $c_{31}^{1}+v_{31}^{1}=2$, and $c_{21}^{1}+v_{21}^{1}=1$. The number of carriages that need to be attached to the train when the train passes through station 4 should be $c_{43}^{1}+v_{43}^{1}+c_{42}^{1}+v_{42}^{1}+c_{41}^{1}+v_{41}^{1}=6$. When the train passes through station 3, $c_{43}^{1}+v_{43}^{1}=2$ carriages are detached from the train and $c_{32}^{1}+v_{32}^{1}+c_{31}^{1}+v_{31}^{1}=6$ carriages are attached to the train. In other words, the number of carriages should be $6-2+6=10$. This can also be calculated as $c_{31}^{1}+v_{31}^{1}+c_{32}^{1}+v_{32}^{1}+c_{41}^{1}+v_{41}^{1}+c_{42}^{1}+v_{42}^{1}=10$. By using the calculation method of formula (2) and setting $\tilde{h}=3$ and $m=4$, the maximum number of carriages when the train passes through station 3 is calculated as $\sum_{i=3}^{4}\sum_{j=1}^{3-1}\left(c_{ij}^{1}+v_{ij}^{1}\right)=10$.

Constraints (3) to (5) indicate that the number of carriages parked at a platform at the same time does not exceed the maximum number of carriages that can be parked at the platform at the same time. When train $k-1$ passes through station j, $\sum_{i=j+1}^{m}c_{ij}^{k-1}$ carriages are detached from train $k-1$ and then parked at the platform of station j. At that time, $\sum_{i=1}^{j-1}c_{ji}^{k}$ carriages have parked at the platform of station j for passengers to board. Therefore, the sum of $\sum_{i=j+1}^{m}c_{ij}^{k-1}$ and $\sum_{i=1}^{j-1}c_{ji}^{k}$ should be smaller than or equal to T_{pla}. Note that trains k and $k+2$ will use the same sub-area.

Constraint (6) ensures the conservation of the number of carriages on each storage line. Fig. 6 demonstrates the meaning of the first case of constraint (6). The other cases of the constraint have similar meanings. As shown in Fig. 6, when train k passes through station j, $\sum_{i=j+1}^{m}v_{ij}^{k}$ vacant carriages will be detached from the train and parked on the storage line. Meanwhile, $\sum_{i=1}^{j-1}v_{ji}^{k}$ carriages will be attached to the train from the storage line. Additionally, the number of carriages that will be detached from the train and parked at the platform $\sum_{i=j+1}^{m}c_{ij}^{k}$ is not necessarily equal to the number of carriages required by the station $\sum_{i=1}^{j-1} c_{ji}^{k+2}$. If $\sum_{i=1}^{j-1} c_{ji}^{k+2}$ is larger than $\sum_{i=j+1}^{m} c_{ij}^{k}$, then the extra $\sum_{i=1}^{j-1} c_{ji}^{k+2} - \sum_{i=j+1}^{m} c_{ij}^{k}$ vacant carriages will be transferred from the storage line and parked at the platform (see the 4th term in the first case of constraint (6)). On the contrary, if $\sum_{i=1}^{j-1} c_{ji}^{k+2}$ is smaller than $\sum_{i=j+1}^{m} c_{ij}^{k}$, the extra $\sum_{i=j+1}^{m} c_{ij}^{k} - \sum_{i=1}^{j-1} c_{ji}^{k+2}$ carriages will become vacant after all passengers inside get off and will be parked on the storage line when train $k+2$ arrives. Similarly, when train k arrives, the extra $\sum_{i=j+1}^{m} c_{ij}^{k-2} - \sum_{i=1}^{j-1} c_{ji}^{k}$ carriages will be parked on the storage line (see the last term in the first case of constraint (6)).

Constraints (7) to (9) ensure that the number of carriages parked on the storage line does not exceed the capacity of a storage line. It is worth mentioning that when the last two trains pass through a station, the tracks near the platform are regarded as a part of the storage line. Constraints (10) and (11) are for the conservation of the number of passengers on platforms when a train arrives. Passengers on a platform not only include the ones who just arrive at the platform but also include the remaining passengers on the platform who have not got on the previous train. Constraint (12) means that the number of carriages required by the system does not exceed the maximum number of carriages in the system. The carriages required by the system include the carriages parked on storage lines and platforms of stations 2 to m when trains 1 and 2 arrive. Constraints (13) and (14) guarantee that the number of carriages is a non-negative integer.

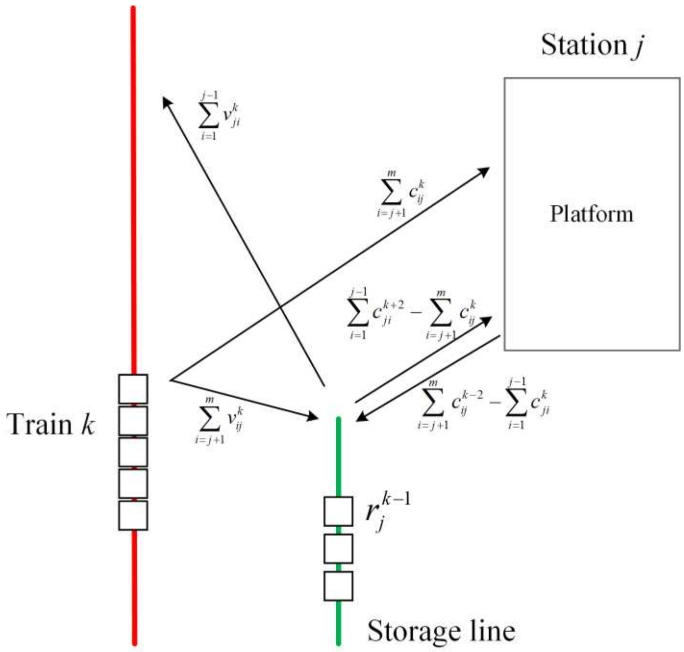


Fig. 6. Schematic diagram of changes in the number of carriages on a storage line.

In addition, there are interferences at the junction of two sub-areas in the double-area system. For example, if there are many carriages parked at sub-area 2 in Fig. 4, these carriages may occupy the space near sub-area 1 when entering the main track, as these carriages need to turn when crossing tracks. Therefore, a certain space needs to be reserved at the junction to reduce the interferences. The model reserves the length of one carriage at the junction as shown in constraints (3) to (5), (8), and (9). In reality, the length of reserved space can be increased by modifying the model to reserve the length of more carriages or taking a smaller T_{pla}-value.

3.2. Model of the traditional URT system

Assumptions (1) and (2) in Section 3.1 are adopted here. Applying those notations summarized in Table 3, we propose the following optimization model of the traditional URT system:

$$\min_{\tilde{C}} \sum_{k\in K}\sum_{(i,j)\in S}\left(\delta\cdot\left(h_{ij}^{k}-x_{ij}^{k}\right)\cdot t+x_{ij}^{k}\cdot\frac{t}{2}\right), \tag{15}$$

$$s.t.\quad \sum_{i=\tilde{h}}^{m}\sum_{j=1}^{\tilde{h}-1}\tilde{c}_{ij}^{k}\le T_{\mathrm{capa}},\ \tilde{h}=2,3,\cdots,m,\ k\in K, \tag{16}$$

$$T_{\mathrm{rem}}^{ik}=\begin{cases}T_{\mathrm{capa}}-\sum_{\tilde{i}=i+1}^{m}\sum_{\tilde{j}=1}^{i-1}\tilde{c}_{\tilde{i}\tilde{j}}^{k}, & i=2,3,\cdots,m-1,\ k=1,2,\cdots k_{\max},\\ T_{\mathrm{capa}}, & i=m,\ k=1,2,\cdots k_{\max},\end{cases} \tag{17}$$

$$\sum_{j=1}^{i-1}\tilde{c}_{ij}^{k}=\min\left\{T_{\mathrm{rem}}^{ik},\sum_{j=1}^{i-1}h_{ij}^{k}\right\},\ k\in K,\ i=2,3,\cdots,m, \tag{18}$$

$$h_{ij}^{1}=x_{ij}^{1},\ (i,j)\in S, \tag{19}$$

$$h_{ij}^{k}=h_{ij}^{k-1}+x_{ij}^{k}-\tilde{c}_{ij}^{k-1},\ (i,j)\in S,\ k=2,3,\cdots,k_{\max}, \tag{20}$$

$$\tilde{c}_{ij}^{k}\le h_{ij}^{k},\ (i,j)\in S,\ k\in K, \tag{21}$$

$$\tilde{c}_{ij}^{k}\in N,\ (i,j)\in S,\ k\in K. \tag{22}$$

The model of the traditional URT system is similar to that of the double-area system. The objective of the model is to minimize the total waiting time of passengers on platforms. In the traditional URT system, the configuration of a train is fixed, so the number of carriages is not involved in the objective function. Constraint (16) ensures that the number of passengers who get on a train does not exceed the capacity of the train. Constraint (17) calculates the remaining capacity of a train when the train is about to arrive at a station. Note that the non-negativity of the remaining capacity of a train can be guaranteed by constraint (16). Constraint (18) ensures the conservation of the number of passengers who get on a train. If there is space in a train, passengers will get on the train. Constraints (19) and (20) ensure the conservation of the number of passengers on platforms when a train arrives. Constraint (21) guarantees that the number of passengers who get on a train does not exceed the number of passengers on platforms. Constraint (22) ensures that the number of passengers who get on trains is a non-negative integer.

3.3. Model linearization and segmentation

The two models are nonlinear models due to the existence of $\min\{\cdot\}$ and $\max\{\cdot\}$ functions. However, all those formulae containing $\min\{\cdot\}$ function can be converted into linear formulae through the following mathematical transformation. Let $\breve{x}$, $\breve{y}$, and $\breve{z}$ be three variables. The formula $\breve{z} = \min\{\breve{x}, \breve{y}\}$ is equivalent to the following constraints:

$$\breve{z} \leq \breve{x}, \tag{23}$$

$$\breve{z} \leq \breve{y}, \tag{24}$$

$$\breve{z} \geq \breve{x} + M\left(u_1 - 1\right), \tag{25}$$

$$\breve{z} \geq \breve{y} + M\left(u_2 - 1\right), \tag{26}$$

$$u_1 + u_2 \geq 1, \tag{27}$$

$$u_1, u_2 \in \{0,1\}, \tag{28}$$

where M is a sufficiently large constant. In this way, the two nonlinear models are converted into linear models. Similarly, all those formulae containing $\max\{\cdot\}$ function can also be converted into linear formulae.

It is worth mentioning that when the DANRT system is applied in practice, the peak hours (such as 2 h) can be divided into multiple small time segments (such as 240 s which is the time span we adopted in Section 4). Passenger demand that cannot be served during a time segment will be passed to the next time segment. As a result, the two models should be run once for each time segment. There are two advantages to dividing into small time segments. First, the running times of the two models are acceptable by doing this. If we do not divide the peak hours into small time segments, the number of decision variables will be extremely large and the running times of the models will be unacceptable. Second, we have assumed that the capacity of the garage in the starting station is large. However, to meet the carriage needs for a unidirectional line throughout the day, the garage in the starting station may need to store thousands of carriages. To avoid this case, the vacant carriages needed in the starting station should be transferred from the garage in the terminal station if we consider a bidirectional line. If we divide peak hours into small time segments, the carriage scheduling problem in the bidirectional line can be modeled more easily. In this way, the models can be solved quickly by advanced commercial solvers.

4. Numerical experiments

In this section, we apply the two models in Section 3 to a Beijing Subway line, the Batong Line, and provide observations and operational characteristics of the DANRT system. The Batong line consists of 12 stations (i.e., $m = 12$) and has no transfer station. We only consider a unidirectional line from Tu Qiao station to Sihui Dong station. Fig. 7 shows a heatmap delineating the OD demands of passengers on the Batong Line on the map from 7:00 to 9:00 on October 11, 2017. We can see that most of the passengers take the terminal station, i.e., Sihui Dong station, as their destination. Note that the passenger demand data were obtained from an Automatic Fare Collection (AFC) system. In fact, the passenger demands obtained from the AFC system are not actual passenger demands due to passenger flow control measures and the limit of the number and response time of entrance gates. Nonetheless, they should provide a sufficiently accurate approximation on the passenger volume.

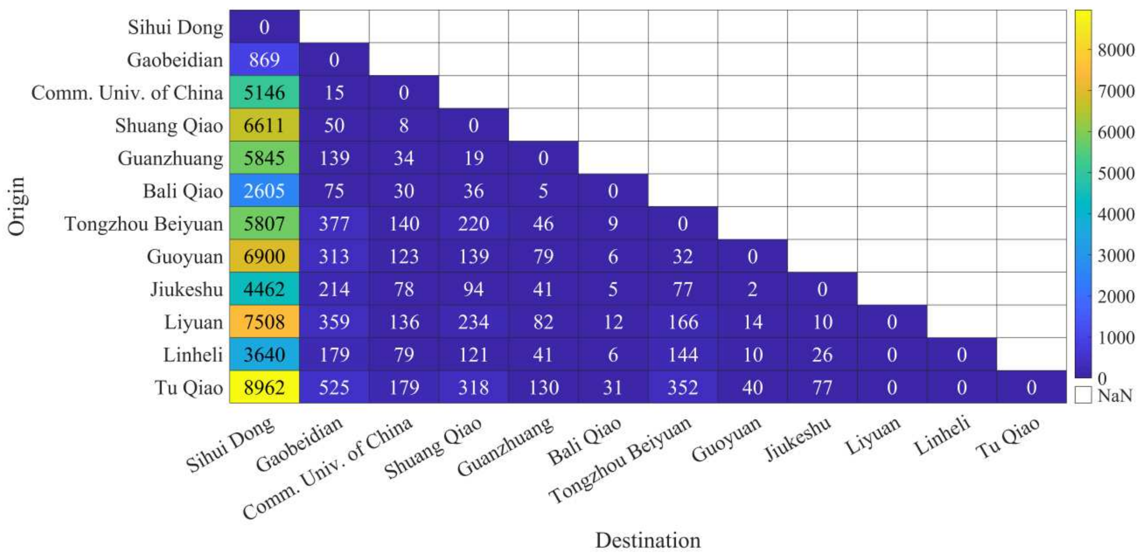


Fig. 7. A heatmap delineating the OD demands of passengers on the Batong Line from 7:00 to 9:00 on October 11, 2017.

According to a set of video recordings taken at the Xierqi station in Beijing Subway, t_s is set to 38. We assume that the train decelerates to 6 m/s when passing a station and the maximum speed, at which a train crosses tracks, is 7 m/s for safe operation of the system according to the *Code for design of metro GB50157-2013*[3] of China. If we use the parameter values in Appendix A, we should set the headway t in the double-area system to be larger than 29.98 s to prevent a train from rear-end collision on main tracks. In addition, we should set the headway t in the double-area system to be larger than 32 s to prevent carriages from rear-end collision on guide tracks if we consider the time taken to pull in and pull out the carriages. In this section, we set the headway t in the model of the double-area system to 40 for safer operation of the system. The headway t in the model of the traditional system should be at least $\left(38\times12+144+120\right)/12=60$ to avoid rear-end collisions if the parameter values in Appendix A are used. The headway t in the model of the traditional system is set to 60. Penalty coefficient δ in the two objective functions is set to 2.

We use the actual OD demands of passengers from 7:00 to 9:00 on October 11, 2017, with an equal proportion increase as the inputs for the two models in Section 3. We use $Redu$ to depict the increase. Specifically, if the passenger demand from station i to station j within two hours is $\hat{x}_{ij}$, then the passenger demand x_{ij}^k from station i to station j between the arrival of train $k-1$ and train k is set to $\text{round}\left(\hat{x}_{ij}/Redu\right)$ in the following experiments. For example, the passenger demand $\hat{x}_{21}$ from Gaobeidian Station to Sihui Dong Station within two hours is 869 as shown in Fig. 7. If we set $Redu=10$, then the passenger demand x_{21}^k ($k\in K$) from Gaobeidian Station to Sihui Dong Station between the arrival of train $k-1$ and train k is set to $\text{round}\left(869/Redu\right)=87$. Each OD demand of passengers is increased by $7200/\left(40\cdot Redu\right)$ times, where 7200 is the number of seconds of two hours and 40 is the headway t in the model of the double-area system. A larger $Redu$ means a smaller demand of passengers. The reasons why we use the OD demands with an increase are as follows. First, the DANRT system is designed for the continuous increase of passenger flow in the future and we want to explore the potential of the DANRT system in improving the travel efficiency of passengers. When we use the actual OD demands of passengers, we cannot explore the potential of the DANRT system because almost all passengers in the two systems can get on the current train. Second, we shorten the headway between successive trains in the traditional system from a few minutes in the real world to 60 s in our experiments. It is inappropriate to use actual OD demands of passengers directly.

The length of a train or a platform in the traditional system is 120 m and the total capacity

[3] https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/201308/20130820_224416.html

T_{capa} of a train in the traditional system is set to 2400. However, the total capacity of a train in the DANRT system should be smaller because no one can stand at the joint of adjacent carriages in the DANRT system. We assume that the gap between two adjacent carriages in the double-area system is 0.2 m. The maximum number of passengers per unit area in a train in the two systems should be equal, and thus the capacity of a carriage c_{capa} is $2400/T_{pla} - 0.2 \times 2400 \div 120$, i.e., $c_{capa} = 2400/T_{pla} - 4$. We consider a time span of 240 s. Therefore, there are $k_{max} = 6$ trains running in the double-area system and $k_{max} = 4$ trains running in the traditional system. Fig. 8 presents a schematic diagram of the cumulative arrival and departure curves of passengers and time intervals in the two systems when the aforementioned settings are applied.

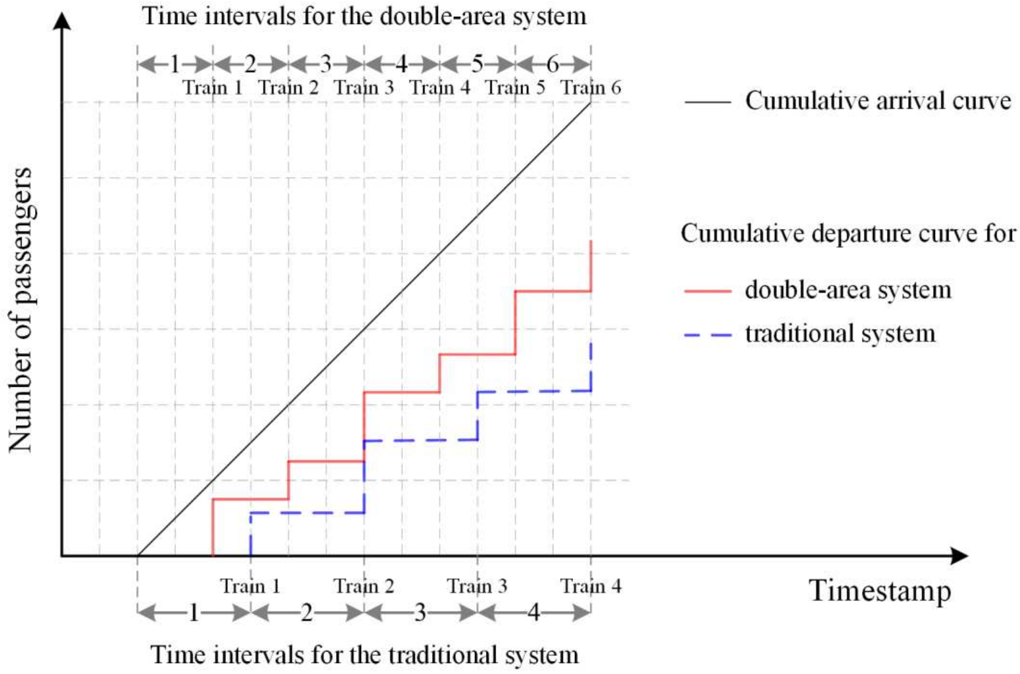


Fig. 8. A schematic diagram of the cumulative arrival and departure curves of passengers and time intervals in the two systems.

We want to explore the potential of the DANRT system and thus the maximum number of carriages in the system γ is set to be a large number of 500. It is worth mentioning that in the following numerical experiments, the maximum number of carriages required by the DANRT system is 299.

We use Matlab 2019a, Yalmip, and Gurobi 9.5.1 to solve the two models on a Windows PC with an Intel i7 2.30 GHz CPU with 16 GB RAM. The numbers of variables of the two models are 3087 and 1078, respectively. The numbers of constraints of the two models are

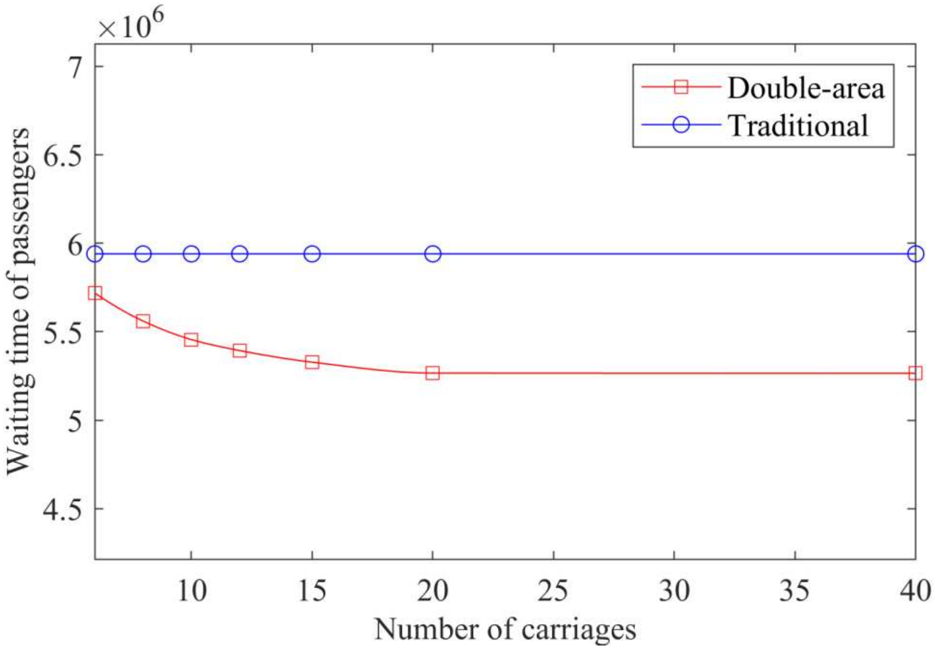
4848 and 2218, respectively. The average running times of the two models are 13.58 s and 2.14 s, respectively.

We investigate the effects of the length of each carriage and the capacity of each storage line on the travel efficiency of passengers in the double-area system when we use the actual OD distribution in Sections 4.1 and 4.2. Subsequently, we explore the effect of different OD distributions of passengers on the travel efficiency of passengers in the two systems in Section 4.3. Meanwhile, we compare the performances of the two systems. Finally, we further investigate the effects of objective functions on the performance of the double-area system and carriage scheduling schemes in Section 4.4.

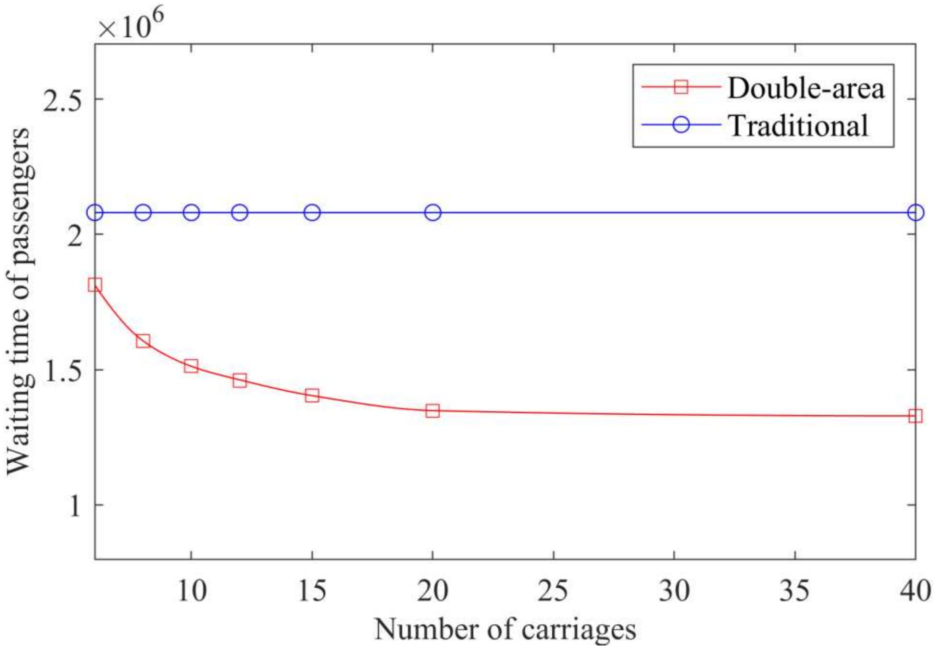
4.1. Effects of the length of each carriage

Since the length of each carriage does not appear in our models in Section 3, we use the maximum number T_{pla} of carriages that can be parked at the platform at the same time to reflect the length of a carriage. Specifically, the length of a carriage is $120/T_{\mathrm{pla}} - 0.2$ m, where 120 m is the length of a platform and 0.2 m is the length of the gap between two adjacent carriages.

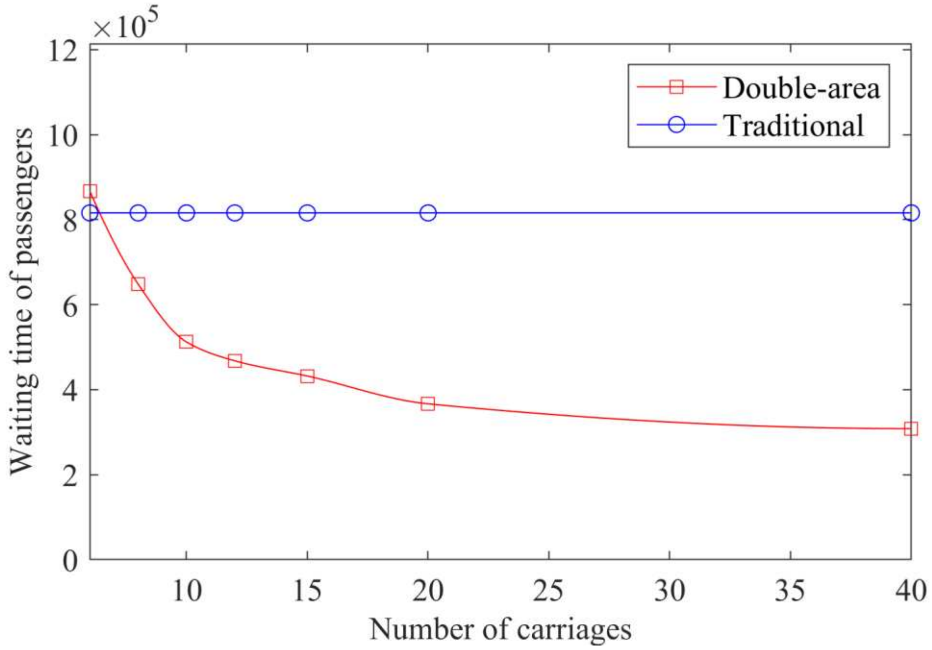
We first set the train length to be equal to the platform length, which is consistent with reality. In other words, the maximum number of carriages a train can load is equal to the maximum number of carriages that can be parked at the platform at the same time, i.e., $T_{\mathrm{tra}} = T_{\mathrm{pla}}$. Additionally, the capacity n of a storage line is set to 0. Fig. 9 shows the relation of the waiting time of passengers (unit: s) to the maximum number T_{pla} ($=6$, 8, 10, 12, 15, 20, or 40) of carriages that can be parked at the platform at the same time when $Redu = 10$, 20, or 30. Note that $Redu = 10$, 20, or 30 means the average inbound passenger flow is 14.53, 7.28, or 4.85 people/s, respectively. In fact, we can hardly find a station with a passenger flow of 14.53 people/s. Additionally, T_{tra} in the double-area system cannot be equal to 40 to prevent trains from rear-end collisions on main tracks. We only use these two extreme values to explore the characteristics of the system. One can see that the total waiting time of passengers in the double-area system is shorter than that in the traditional system in most cases. This means the double-area system performs better in reducing the waiting time of passengers due to the shorter headway between successive trains. When $Redu$ takes 30 and T_{pla} takes 6, the waiting time of passengers in the double-area system is longer than that in the traditional system. The reason is that the model reserves the length of one carriage at the junction of two sub-areas as shown in constraints (3) to (5), (8), and (9). When T_{pla} takes 6, the model reserves 20 m at the junction of two sub-areas, which results in a huge waste of platform space.



(a)



(b)

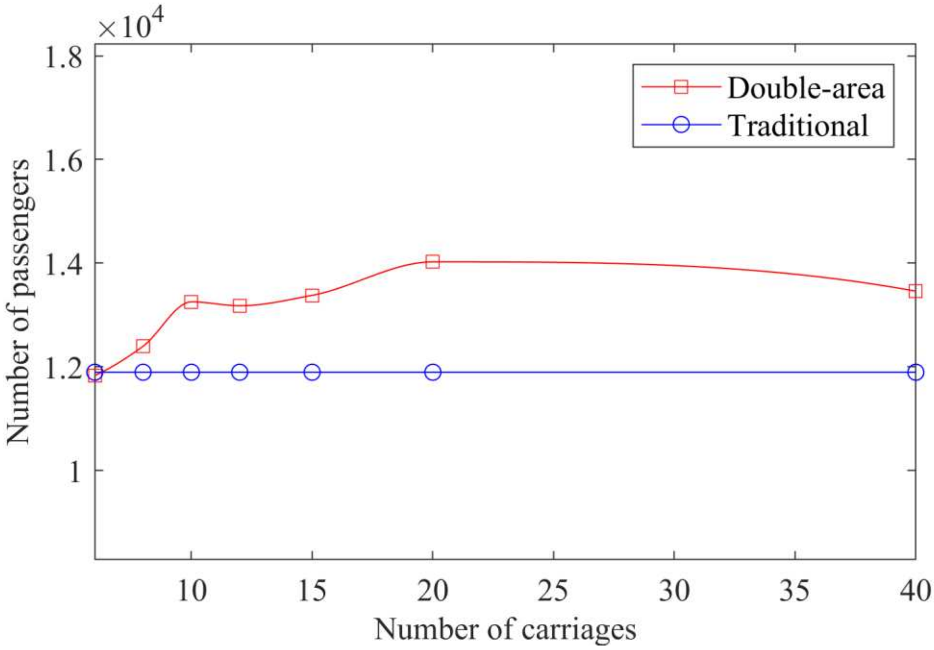


(c)

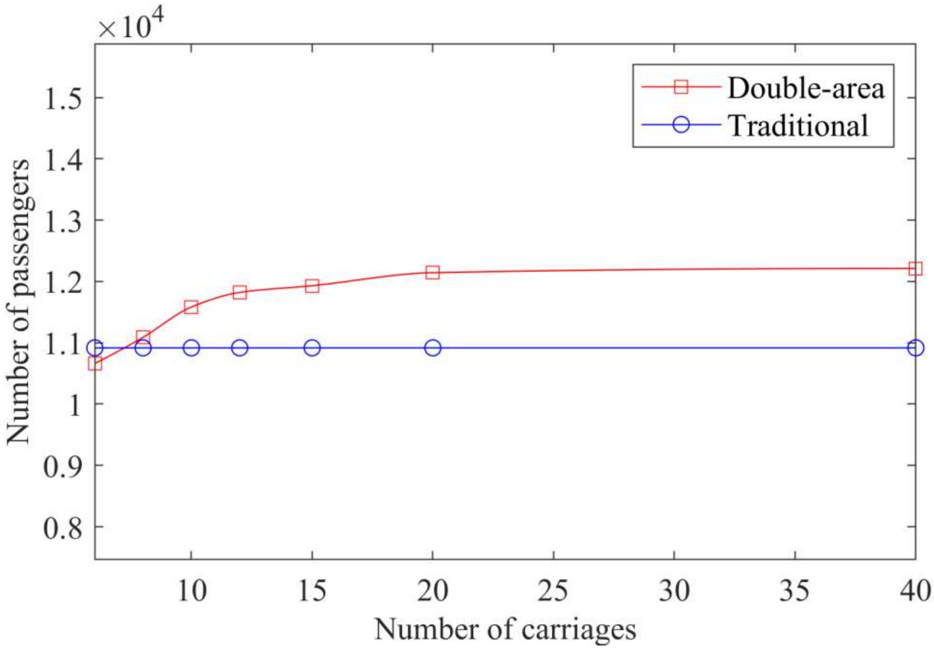
Fig. 9. Relation of the total waiting time of passengers (unit: s) to the maximum number T_{pla} of carriages that can be parked at the platform at the same time when $Redu$ takes (a) 10, (b) 20, or (c) 30.

Meanwhile, the total waiting time of passengers in the double-area system decreases as the increase of T_{pla}. This is because the capacity of a carriage is more likely to be properly utilized when T_{pla} is larger. Consequently, a larger T_{pla} leads to higher capacity utilization of a carriage, and thus the total waiting time of passengers becomes shorter accordingly.

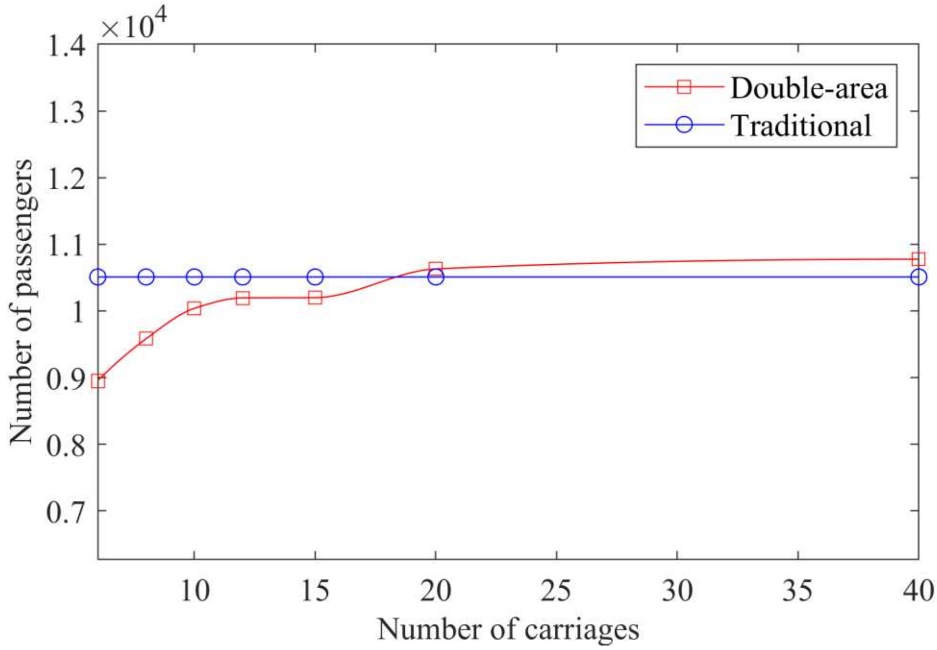
When we minimize the total waiting time of passengers, we obtain a set of optimal carriage scheduling schemes. To understand the DANRT system better, we give the numbers of passengers who complete their travel and the average times saved by each passenger on trains when applying those optimal schemes. Fig. 10 delineates the relation of the number of passengers who complete their travel to the number T_{pla} $(=6$, 8, 10, 12, 15, 20, or 40) of carriages that can be parked at the platform at the same time when $Redu=10$, 20, or 30 in the case of applying the aforementioned optimal schemes. One can see that the performance of the double-area system is better in most cases. The fluctuation of the curve is because the objective function is not the number of passengers who complete the travel. The results obtained by maximizing the number of passengers who complete the travel can be seen from Fig. 16.


×10^4
1.8
1.6
1.4
1.2
1
Number of passengers
Double-area
Traditional
10
15
20
25
30
35
40
Number of carriages

(a)


×10^4
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
Number of passengers
Double-area
Traditional
10
15
20
25
30
35
40
Number of carriages

(b)



(c)

Fig. 10. Relation of the number of passengers who complete their travel to the number T_{pla} of carriages that can be parked at the platform at the same time when *Redu* takes (a) 10, (b) 20, or (c) 30.

The numbers of carriages used in the numerical examples are shown in Fig. 11. One can see that the maximum number of carriages used in the system does not exceed 300 even if T_{pla} takes the extreme value of 40. As the increase of T_{pla}, the number of carriages used in the system increases. In addition, as the decrease of passenger demand, the number of carriages used in the system decreases.

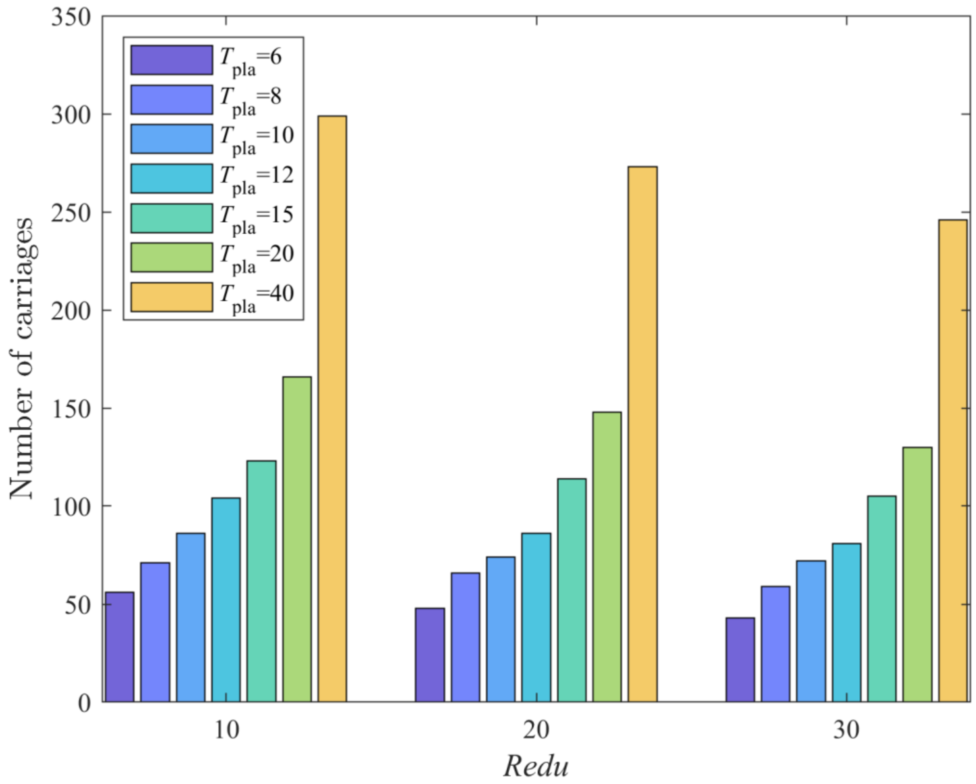


Fig. 11. Numbers of carriages used in the numerical example.

We then give the average times saved by each passenger on trains when applying the aforementioned optimal schemes. In Section 2.2, we have illustrated that a passenger on a train in a DANRT system can save a certain amount of time t_{save} when he/she passes through each station because trains keep running all the way. The average time saved by each passenger on trains (unit: s) can be then calculated as

$$\overline{t} = \frac{\sum_{k\in K}\sum_{(i,j)\in S}\min\left\{c_{ij}^{k}\cdot c_{\text{capa}}, h_{ij}^{k}\right\}\cdot(i-j-1)\cdot t_{save}}{\sum_{k\in K}\sum_{(i,j)\in S}\min\left\{c_{ij}^{k}\cdot c_{\text{capa}}, h_{ij}^{k}\right\}}. \tag{29}$$

Some passengers cannot complete their travel within a time span of $k_{\max}\cdot t_{s}$. Therefore, we only calculate the saved time of passengers who complete their travel. It is worth mentioning that even if we use the same demands of passengers for a certain period of time, the two sets of passengers who complete their travel in the double-area and traditional systems are different. As a result, we cannot compare how much time a person would save if he/she chose the other system. Therefore, we calculate the average time saved by each passenger on trains in the double-area system. We assume that the train decelerates to 6 m/s when passing a station. t_{save} is set to 26.63, 25.01, 23.56, 22.19, 20.23, 17.06, or 4.81 when T_{tra} takes 6, 8, 10, 12, 15, 20, or 40 respectively according to Appendix A.

Fig. 12 shows the average time saved by each passenger on trains. One can see that the average times saved by each passenger on trains in the double-area system increase as the increase of $Redu$. The average time exceeds 50 s when $Redu = 10$ and reaches about 150 s when $Redu = 30$. (Note that T_{tra} cannot be set to 40 to avoid rear-end collisions.) This means passengers who complete their travel are more likely to save more time on trains when the demands of passengers are small. The reason is as follows. According to the solutions of the models, we find that passengers who have near destinations are more likely to get on a train compared with the ones who have distant destinations. Meanwhile, passengers who have distant destinations can save more time compared with the ones who have near destinations. Consequently, when the demands of passengers are smaller, more passengers who have distant destinations can get on a train.

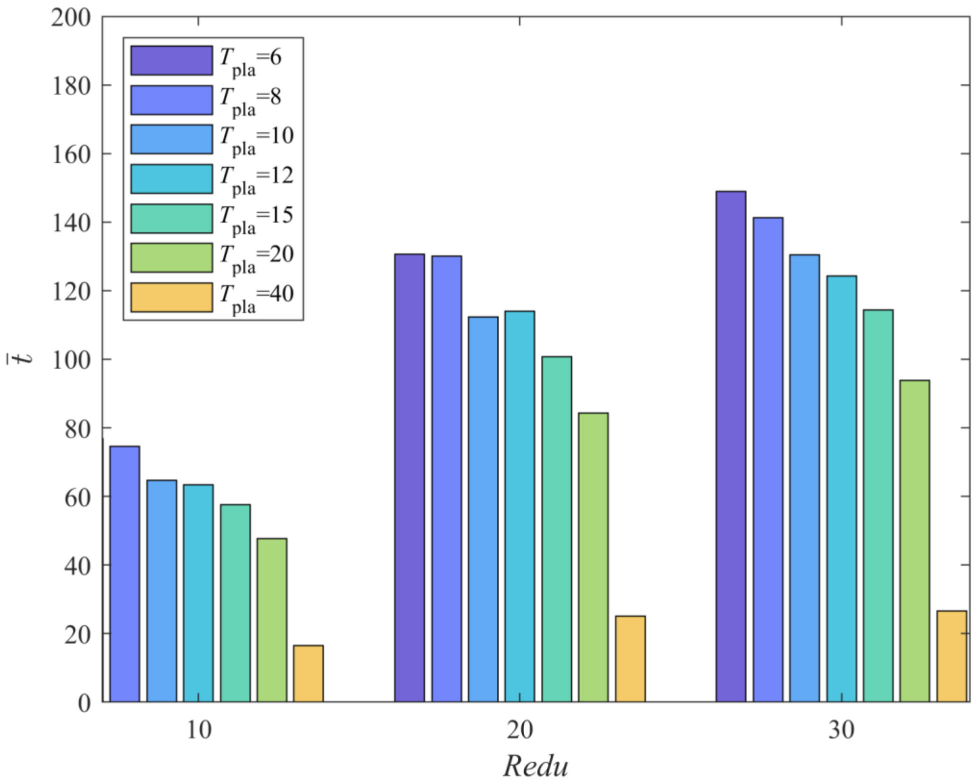


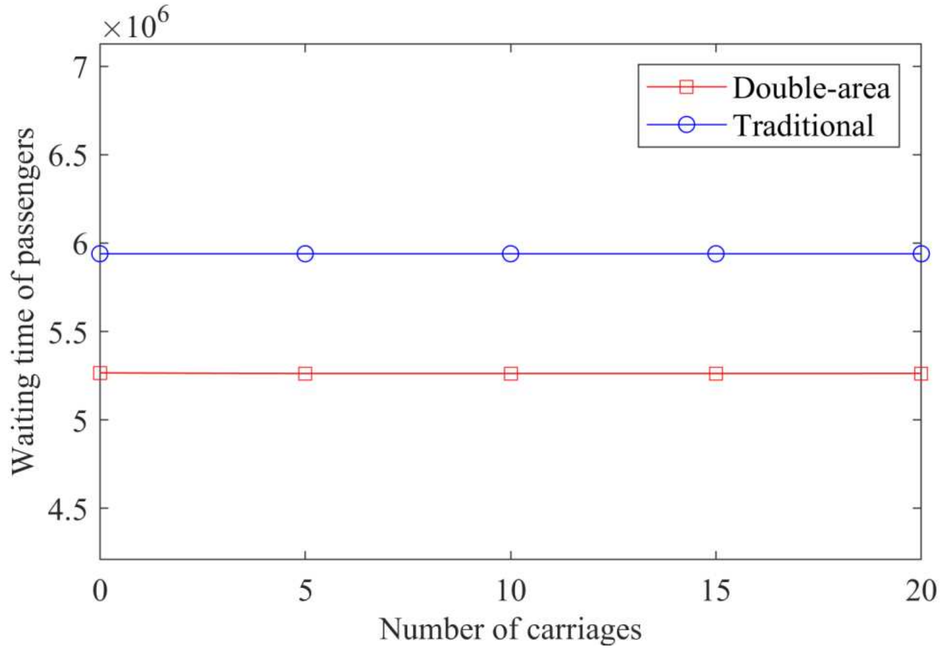
Fig. 12. Average time saved by each passenger on trains (unit: s) in the double-area system.

It is worth mentioning that it takes 29 minutes for a train to complete the entire journey on the Batong Line. Therefore, we conclude that passengers in DANRT can save about $50/(29\times 60) = 2.9\%$ to $150/(29\times 60) = 8.6\%$ of travel time compared with the traditional system.

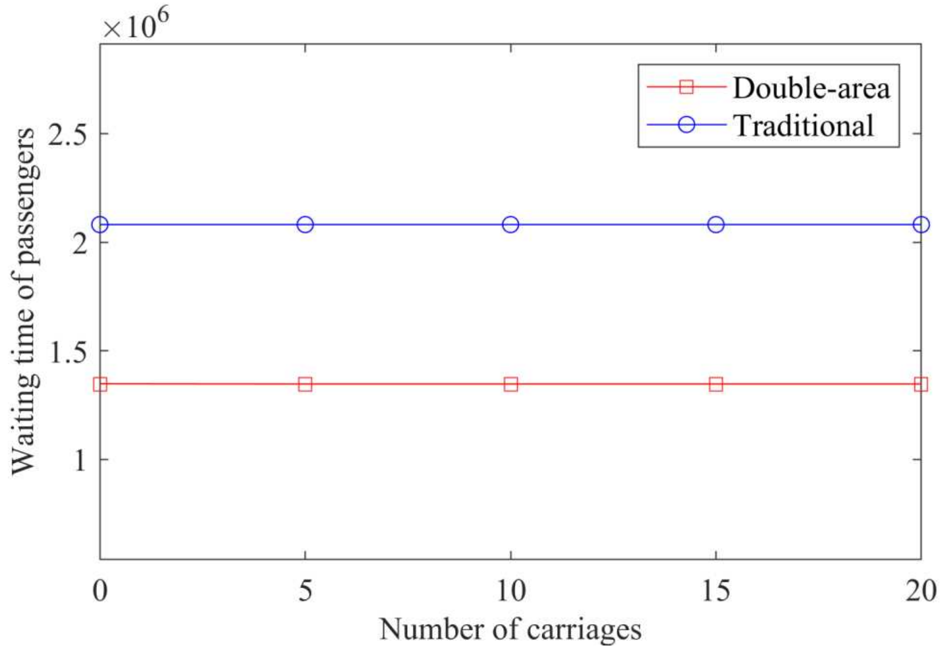
4.2. Effect of the capacity of storage lines

We then investigate the effect of the capacity n of storage lines on the waiting time of passengers on platforms. We set $T_{\text{pla}} = T_{\text{tra}} = 20$. Fig. 13 delineates the relation of the waiting time of passengers to the capacity n ($=0$, 5, 10, 15, or 20) of storage lines when $Redu = 10$, 20, or 30. One can see that the waiting times of passengers remain almost unchanged as the increase of n in all cases. Almost all vacant carriages needed by stations to accommodate passengers can be supplied by trains, and thus the waiting times of passengers cannot be reduced by increasing n.

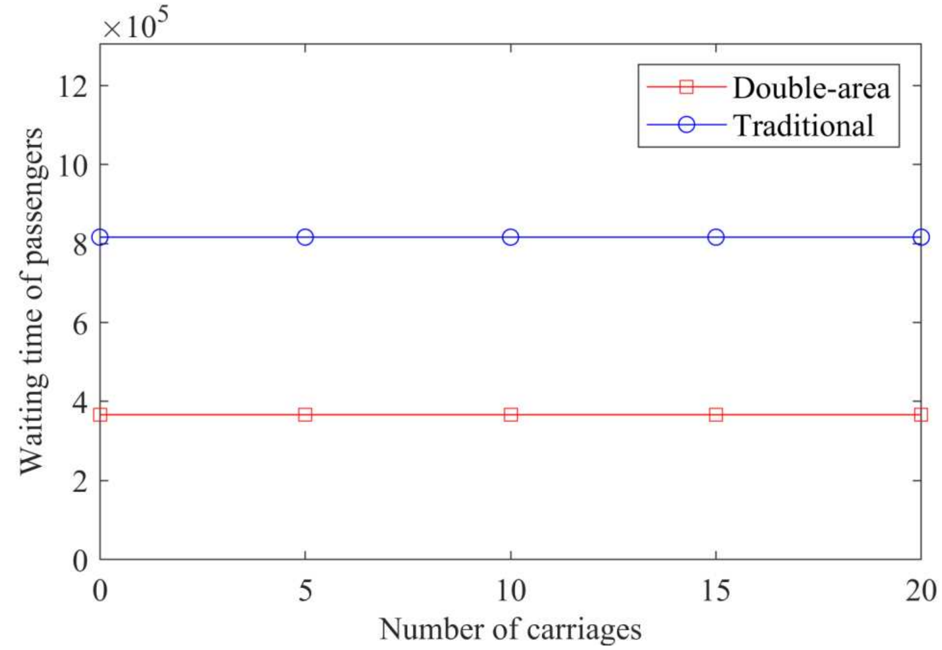
These results mean that the double-area system hardly performs better by increasing the capacity of a storage line. The benefit of building a storage line in each station thus might be marginal. In other words, the storage line in each station is unnecessary in the double-area system when we use the OD distribution of passengers on the Batong Line of Beijing Subway as model input.



(a)



(b)



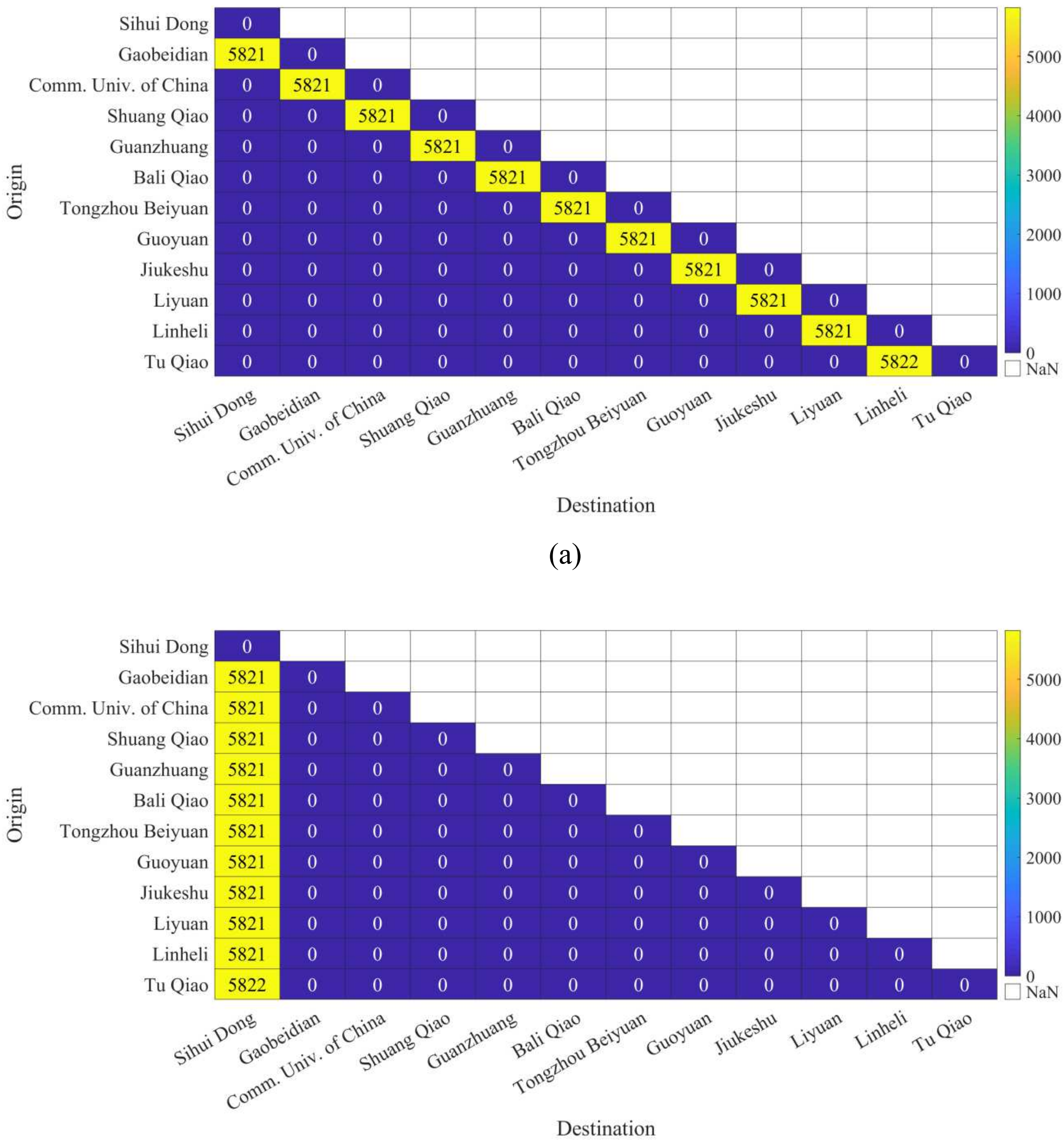
(c)

Fig. 13. Relation of the waiting time of passengers (unit: s) to the capacity n of storage lines when $Redu$ takes (a) 10, (b) 20, or (c) 30.

4.3. Effect of the OD distribution of passengers

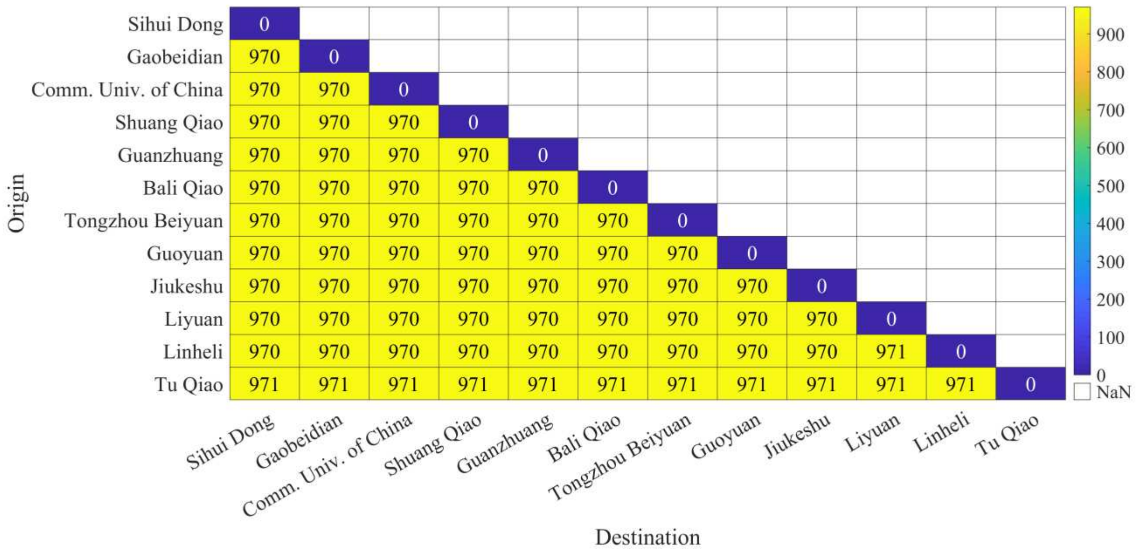
In the above two subsections, we use the actual OD distribution as the input. We want to explore whether OD distribution affects the results. In this subsection, we investigate the effect of the OD distribution of passengers on the waiting time of passengers on platforms.

We generate three typical OD distributions of passengers. The first is that each passenger gets on a train at the current station and then gets off at the next station. The second is that each passenger gets off at the terminal station. The third is a balanced demand of passengers. For the sake of clarity, we call the first OD distribution of passengers **OD distribution 1**, the second OD distribution of passengers **OD distribution 2**, and the third OD distribution of passengers **OD distribution 3**. Meanwhile, the number of arrivals at each station to each destination is set to be equal for simplicity as shown in Fig. 14. To compare the results in this subsection with those in Section 4.1, we set the total number of passengers entering all stations within a certain time to be equal to that in Section 4.1 when *Redu* takes each value.



(a)

(b)

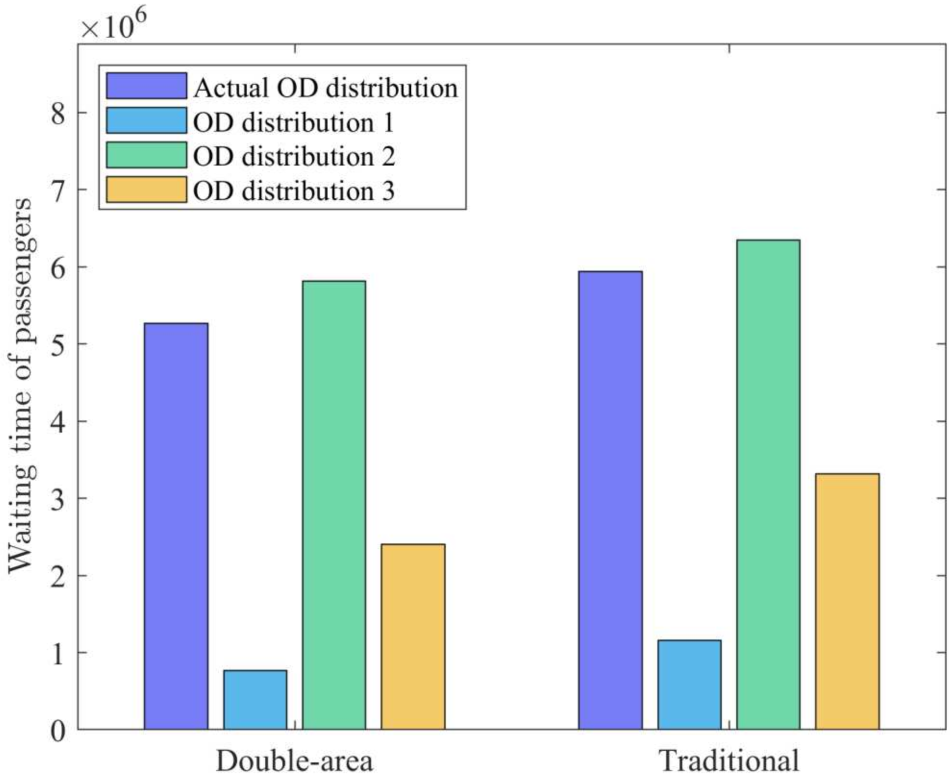


(c)

Fig. 14. Three heatmaps delineating the generated OD demands of passengers. (a) for OD distribution 1, (b) for OD distribution 2, and (c) for OD distribution 3.

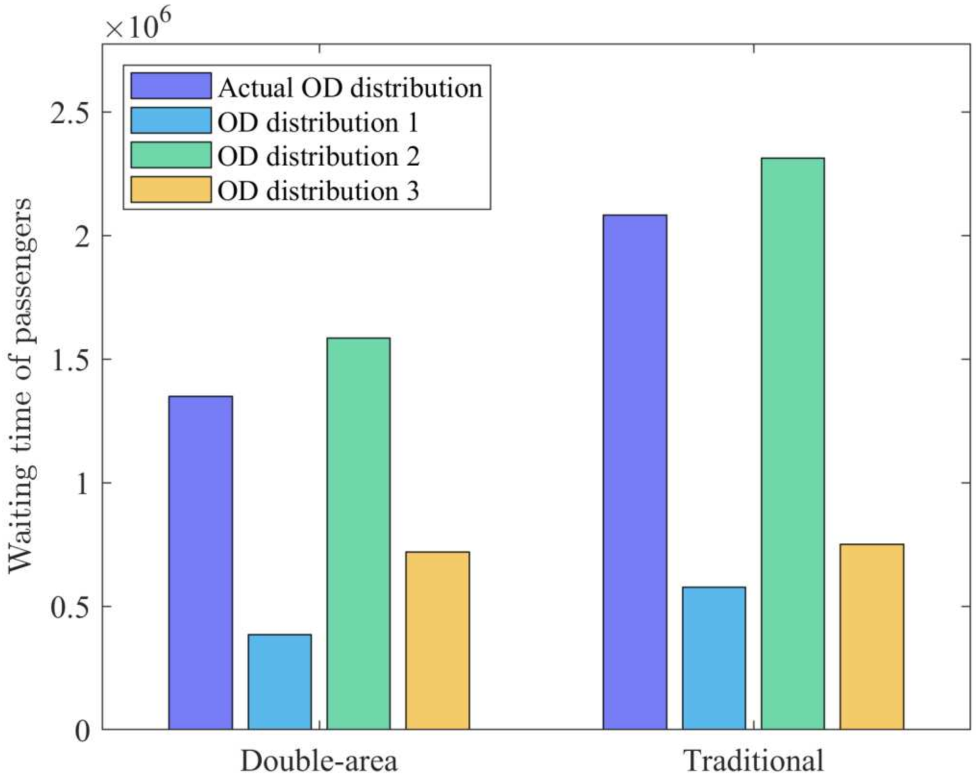
We set $T_{\text{tra}} = T_{\text{pla}} = 20$ and $n = 0$ in the model of the double-area system. Fig. 15 shows the effect of the OD distribution of passengers on the waiting time of passengers when $Redu = 10$, 20, or 30. Overall, the performance of the double-area system is better even if the OD distribution varies in most cases. In the case of $Redu = 30$, however, the performance of the traditional system is better when OD distribution 3 is used. In addition, when OD distribution 3 is used, the waiting time of passengers is the longest in the double-area system in the case of $Redu = 30$. The reason is that in this case, passenger demand to any destination is significantly smaller than the capacity of a carriage in the DANRT system. Thus, the capacity of each carriage cannot be properly utilized. In other words, a DANRT system is more suitable for managing large demand, especially for unbalanced demand. Note that in cases where a DANRT system is not suitable, a DANRT system can be transformed into the traditional one, as mentioned in Section 2.2.

In addition, we observe that when OD distribution 1 is used as the input for the two models, the waiting time of passengers is the shortest. This is understandable because when a train/carriage is parked at each station, all passengers inside will get off, which leads to a large remaining capacity of the train. On the contrary, when OD distribution 2 is used as the input for the two models, the waiting time of passengers is the longest in most of the cases. The reason is that once passengers get on a train, they do not get off until the terminal station, which results in a small remaining capacity of the train when the train is parked at each station.



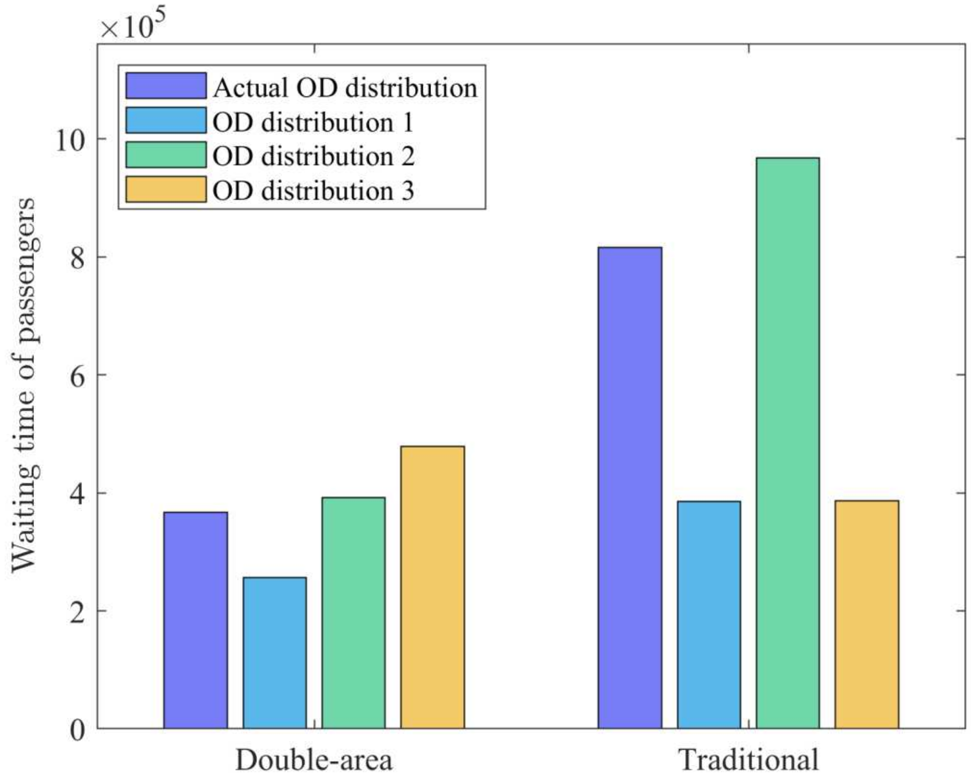
×10^6
Actual OD distribution
OD distribution 1
OD distribution 2
OD distribution 3
Waiting time of passengers
0
1
2
3
4
5
6
7
8
Double-area
Traditional

(a)



×10^6
Actual OD distribution
OD distribution 1
OD distribution 2
OD distribution 3
Waiting time of passengers
2.5
2
1.5
1
0.5
0
Double-area
Traditional

(b)



×10^5
Actual OD distribution
OD distribution 1
OD distribution 2
OD distribution 3
Waiting time of passengers
10
8
6
4
2
0
Double-area
Traditional

(c)

Fig. 15. Effect of the OD distribution of passengers on the waiting time of passengers (unit: s) when *Redu* takes (a) 10, (b) 20, or (c) 30.

We can conclude that different OD distributions/demands have different effects on the remaining capacity of a train and the capacity utilization of a carriage. In addition to the length of carriages (see Section 4.1), the capacity utilization of carriages in the DANRT system can also be affected by OD demands of passengers. Moreover, the DANRT system is more suitable for managing large demand, especially for unbalanced demand.

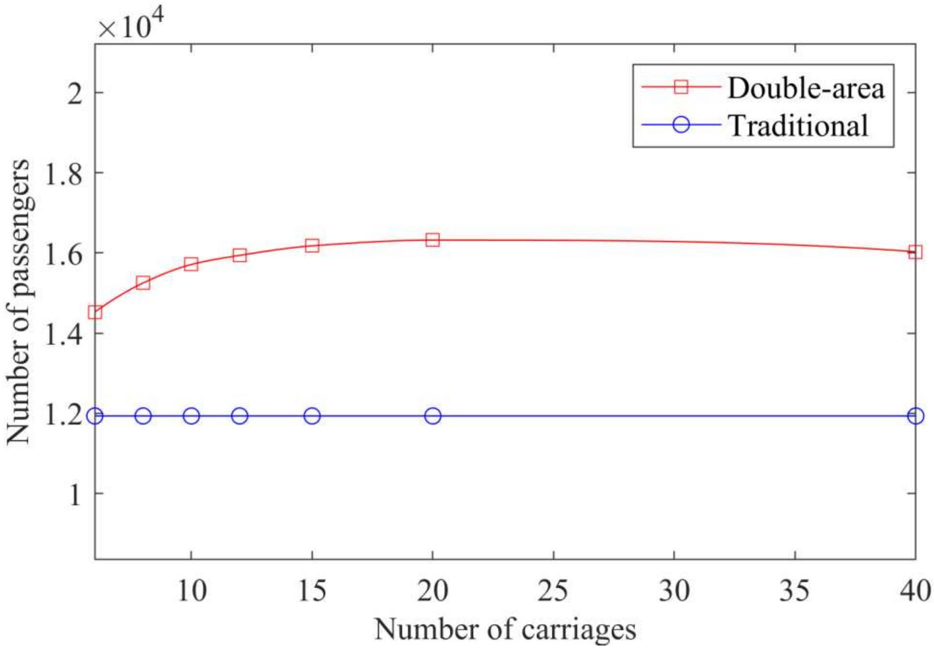
4.4. Effects of objective function

To further explore the characteristics of the DANRT system, this subsection tries different objective functions to verify the system performances.

First, we use the number of passengers completing the travel, i.e.,

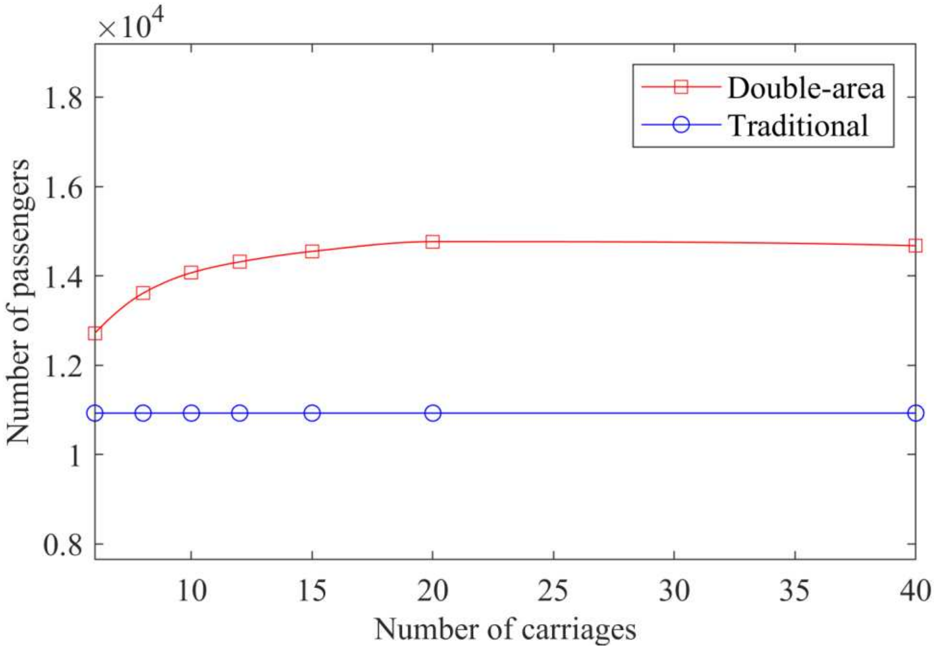
$$\sum_{k\in K}\sum_{(i,j)\in S}\min\left\{c_{ij}^{k}\cdot c_{\text{capa}},h_{ij}^{k}\right\},$$

as the objective function. Fig. 16 shows the relation of the number of passengers who complete their travel to T_{pla} ($=6$, 8, 10, 12, 15, 20, or 40) when maximizing the number of passengers who complete their travel. The other parameter values are the same as those in Section 4.1. One can see that, different from the results shown in Fig. 10, the performance of the double-area system is better in each case. In addition, most curves corresponding to the DANRT system increase first and then decrease. The reason is as follows. On one hand, a larger T_{pla} leads to a smaller capacity of a train due to the existence of the gap between two adjacent carriages. As a result, the number of passengers completing the travel becomes smaller accordingly. On the other hand, the capacity of a carriage is more likely to be properly utilized when T_{pla} is larger as mentioned in Section 4.1. Consequently, a larger T_{pla} leads to higher capacity utilization of a carriage, and thus the number of passengers completing the travel becomes larger accordingly. The two factors affect the number of passengers completing the travel together and cause the number of passengers completing the travel to become larger first and then smaller. In other words, there is a tradeoff between the capacity of a train and the capacity utilization of a carriage when we change the length of each carriage. Moreover, the number of passengers completing the travel is largest when $T_{\text{pla}}=20$.



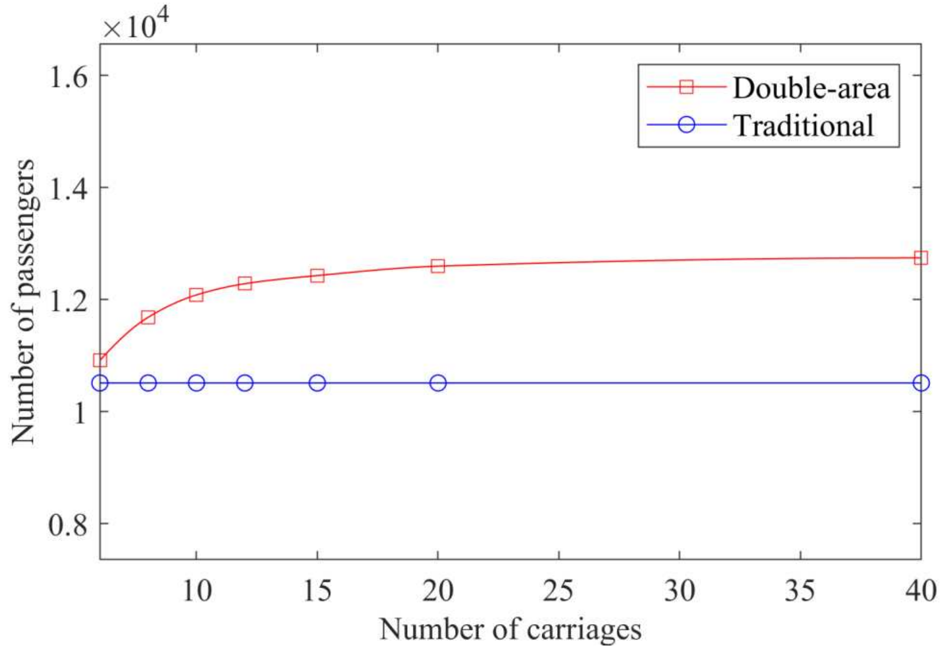
×10^4
Double-area
Traditional
2
1.8
1.6
1.4
1.2
1
Number of passengers
10
15
20
25
30
35
40
Number of carriages

(a)



×10^4
Double-area
Traditional
1.8
1.6
1.4
1.2
1
0.8
Number of passengers
10
15
20
25
30
35
40
Number of carriages

(b)



(c)

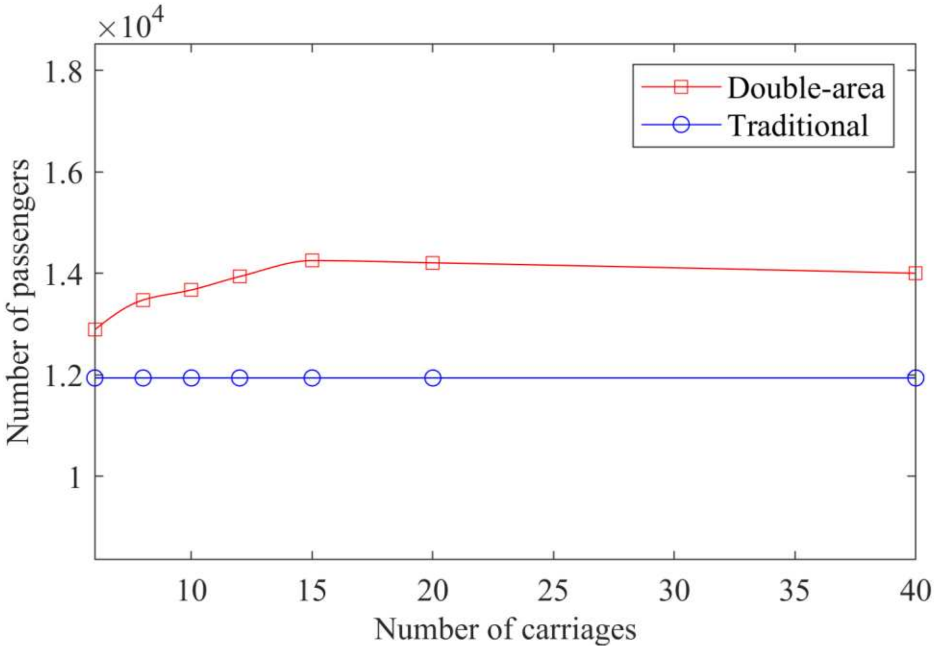
Fig. 16. Relation of the number of passengers who complete their travel to the number T_{pla} of carriages that can be parked at the platform at the same time in the case that *Redu* takes (a) 10, (b) 20, or (c) 30 when maximizing the number of passengers who complete their travel.

However, the number of passengers completing the travel becomes larger as the increase of T_{pla} in the double-area system when *Redu* takes 30. The reason is that, in the double-area system, the capacity of trains is large. In this case, the main reason why passengers cannot get on a train is that there are not enough carriages to their destinations. Therefore, increasing T_{pla} will enable more passengers to complete their travel compared with increasing the capacity of a train through decreasing T_{pla}.

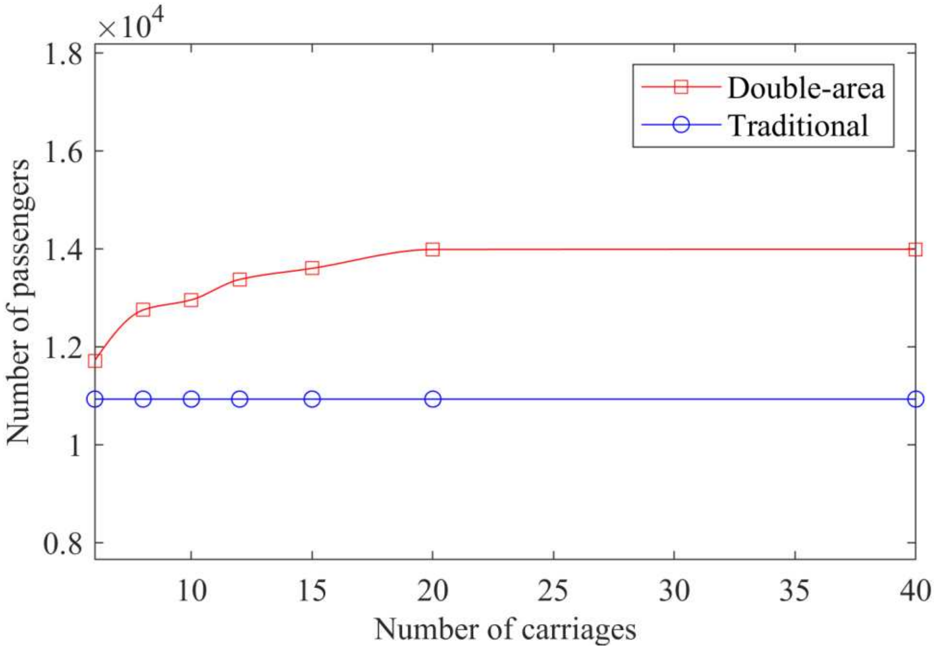
We introduced another possible objective function in Section 3, i.e., the time saved by passengers on trains, which can be calculated as

$$\sum_{k\in K}\sum_{(i,j)\in S}\min\left\{c_{ij}^k\cdot c_{\text{capa}},h_{ij}^k\right\}\cdot(i-j-1)\cdot t_{save}\ .$$

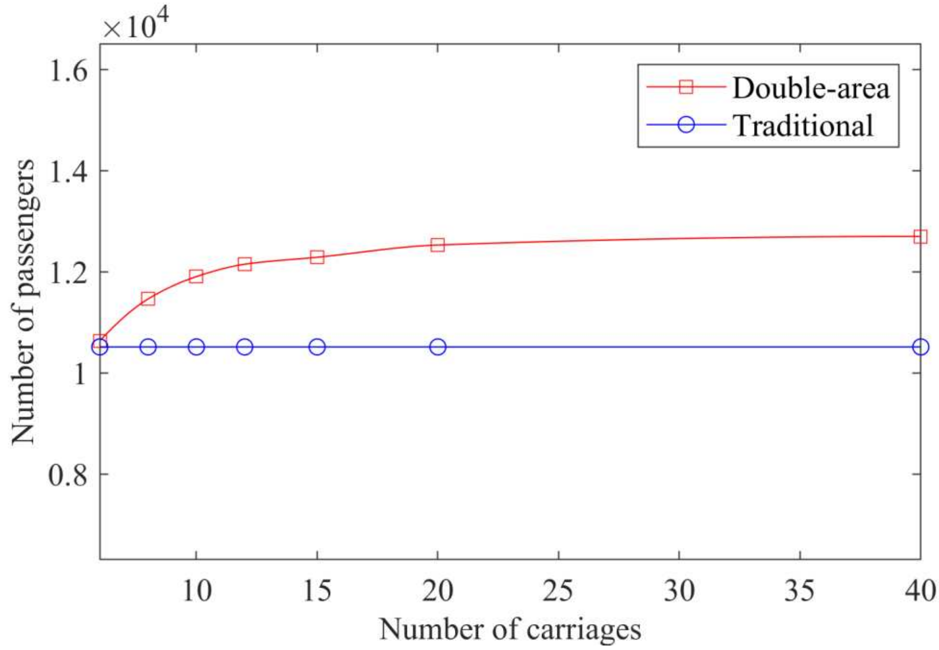
We only need to compare the number of people completing the travel in the double-area system with that in the traditional system. Because the saved time of each passenger on a train is a positive number when we use the parameter values in Appendix A. If the number of passengers completing the travel in the double-area system is larger than that in the traditional system, this indicates that the double-area system is better. Fig. 17 shows the relation of the number of passengers who complete their travel to T_{pla} $(=6$, 8, 10, 12, 15, 20, or 40) when maximizing the saved time of passengers on trains. One can see that the results are similar to those shown in Fig. 16. It is worth mentioning that we also record the waiting time of passengers, and the results are similar to those shown in Fig. 9. In other words, the performance of the double-area system is better than that of the traditional system when we use the waiting time of passengers, the capacity of the system, or the saved time of passengers on trains as the indicator.



(a)



(b)

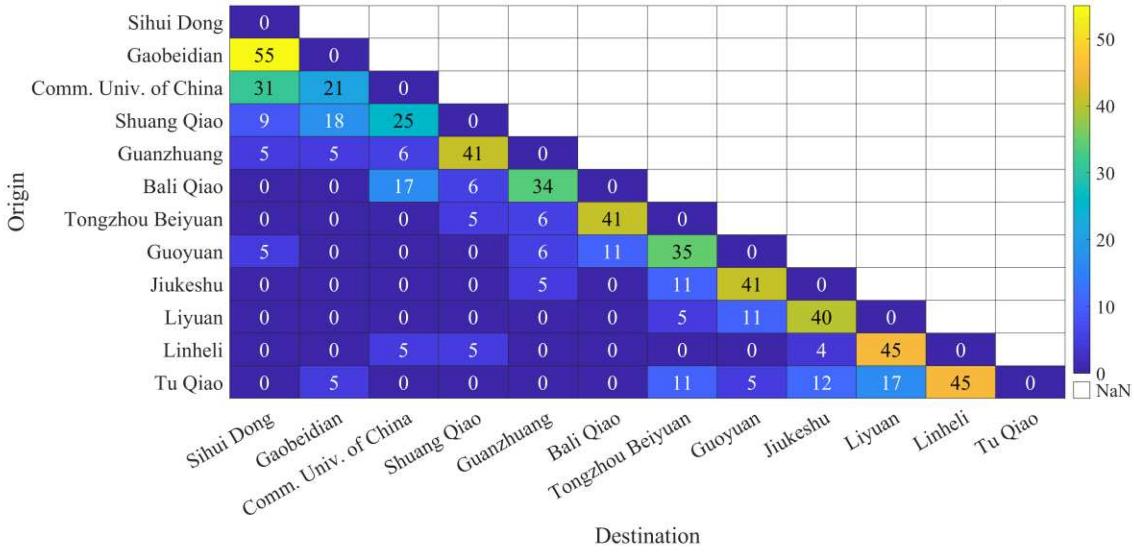


(c)

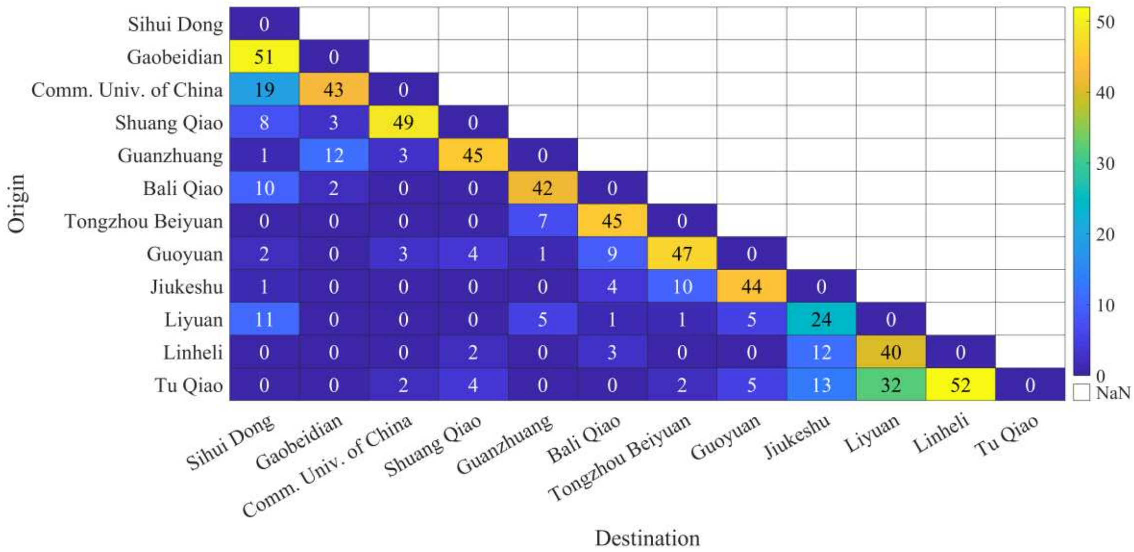
Fig. 17. Relation of the number of passengers who complete their travel to the number T_{pla} of carriages that can be parked at the platform at the same time in the case that $Redu$ takes (a) 10, (b) 20, or (c) 30 when maximizing the saved time of passengers on trains.

We then demonstrate the differences in carriage scheduling schemes when using three different objective functions, i.e., the waiting time of passengers, the capacity of the system, and the saved time of passengers on trains. When the OD distribution of passengers shown in Fig. 7 is used as model input, the differences in carriage scheduling schemes are not significant because most of the passengers take the terminal station as their destination, and most carriages are used to meet the demands of these passengers. Therefore, we use a more balanced OD distribution of passengers shown in Fig. 14(c) as model input. In addition, we set $Redu = 1$ to analyze the carriage scheduling scheme when passenger demands far exceed the capacity of the system. Otherwise, almost all passenger demands will be met and the differences in carriage scheduling schemes are not significant.

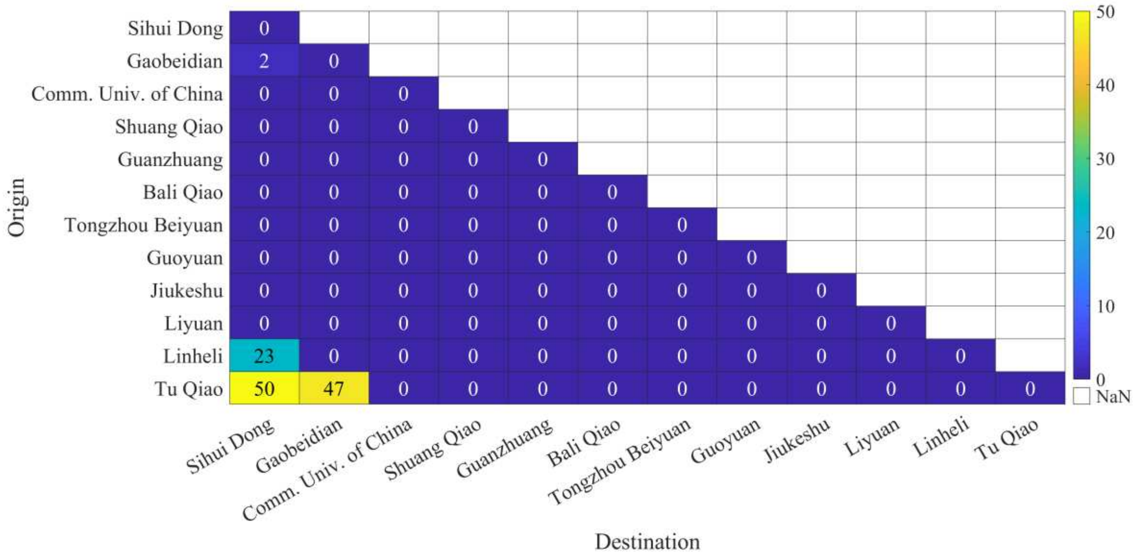
Fig. 18 shows the total number of carriages corresponding to each OD within the focused time span when using the three indicators respectively. One can see that when using the waiting time of passengers or the capacity of the system as the objective function, passengers who intend to go to nearer stations are more likely to get on a train. The two carriage scheduling schemes are similar. However, if the saved time of passengers on trains is used as the objective function, passengers who intend to go to further stations are more likely to get on a train. When using the waiting time of passengers or the capacity of the system as the objective function, the distribution of carriages is more even. By giving the waiting time or the number of passengers who have waited longer a higher weight, the distribution of carriages can become more even. In other words, managers can solve some fairness issues to some extent more easily by using one of these two objective functions and adjusting this weight.



(a)



(b)



(c)

Fig. 18. The total number of carriages corresponding to each OD within the focused time span when using (a) the waiting time of passengers, (b) the capacity of the system, or (c) the saved time of passengers on trains as the indicator.

The following characteristics of the DANRT system can be concluded. Using different objective functions may result in different carriage scheduling schemes. However, the performance of the double-area system is better in reducing the waiting time of passengers on platforms, increasing the number of passengers who complete their travel, and saving the time of passengers on trains due to the shorter headway. Similarly, it can be inferred that further reducing the headway can improve system performance under the premise of preventing trains from rear-end collisions. However, according to Appendix A, it is difficult to further shorten the headway. Therefore, it is enough to divide a platform into two sub-areas, and there is no need to divide the platform into three or more sub-areas.

We also find that the storage line in each station is unnecessary in the double-area system when we use the OD distribution of passengers on the Batong Line of Beijing Subway as model input. It is worth explaining why we still consider storage lines in the DANRT system. First, we only use one OD distribution of passengers as model input. When other OD distributions are used as model inputs, storage lines may be necessary. Second, if there is a storage line in a station, the number of carriages to be transferred will be reduced. In other words, storage lines can help reduce energy consumption, and there is a tradeoff between the construction cost of storage lines and the energy consumption of the system. Therefore, storage lines may be necessary when we consider the construction cost of storage lines and the energy consumption of the system. We will explore this problem in future studies.

There is a tradeoff between the capacity of a train and the capacity utilization of a carriage when we change the length of each carriage. Consequently, we should determine the parameters in the models according to the OD demands of passengers. Moreover, different OD distributions/demands have different effects on the remaining capacity of a train and the capacity utilization of a carriage. This indicates that some traffic demand management measures can be used in the DANRT system for a higher travel efficiency.

5. Conclusions

This paper presents the proposal of a DANRT system and addresses the issue of carriage scheduling within the system using a mathematical programming model. The primary objective of the DANRT system is to enhance the efficiency of passenger travel. By implementing a DANRT system, trains no longer make stops along the entire route, allowing passengers to reach their destinations directly without intermediate delays. The paper outlines a cost-effective design for the DANRT system and provides a comprehensive explanation of train operations within the system. Furthermore, the feasibility of adopting a DANRT system is discussed, along with a thorough comparison of the advantages and disadvantages of DANRT in contrast to traditional URT systems. The findings indicate that constructing a DANRT system would be highly advantageous.

We then formulate the carriage scheduling problem in the DANRT system as an integer programming problem. The number of carriages that will be attached to trains from an upstream station to a downstream station is obtained when the total waiting cost of passengers is minimized. Meanwhile, to illustrate the effectiveness of the DANRT system, we propose a mathematical model to eliminate the effect of passenger arrival sequence on the travel efficiency of passengers in a traditional URT system for comparison. A linearization and segmentation method is proposed to facilitate model solving and the running times of the models are acceptable. To avoid the possible problem of the rapid increase in the number of decision variables in practical applications, we suggest dividing the operation time of a day into small time segments. Efficient algorithms can also be proposed in the future.

Finally, we apply the proposed mathematical models to the Batong Line of Beijing Subway to explore the characteristics of the proposed DANRT system. The main conclusions are summarized as follows. (1) The double-area system performs better than the traditional system. However, the cost is 1.5 times the departure train frequency when the double-area system is used. (2) Passengers in DANRT can save about 2.9% to 8.6% of travel time compared with the traditional system, indicating that the DANRT system have a great potential in improving travel efficiency of passengers. (3) There is a tradeoff between the capacity of a train and the capacity utilization of a carriage. In other words, a moderate size of carriages may result in a higher travel efficiency of passengers. (4) Under certain conditions, the storage line in each station can be unnecessary in the DANRT system, indicating that the construction cost of stations may be lower.

Overall, this paper provides a new perspective to change the traditional system to improve the travel efficiency of passengers. The DANRT system can be applied to URT systems or the intercity trunk lines of urban agglomerations where there are huge passenger demands. The results demonstrate that the DANRT system have a great potential in improving the travel efficiency of passengers. With the growth of the urban population, the development of urban agglomerations, and the advancement of train control technology, this newly designed system may provide a promising direction in URT operations.

This study can be extended in several directions. For example, it is assumed that passenger demands are known. Further studies can consider dynamic and random passenger demands. We only consider a unidirectional line. One can extend the models for a bidirectional line or a network and consider the carriage circulation of such a system. Moreover, energy consumption and construction cost could also be considered.

Appendix A: Possible operational details of trains

This appendix presents several typical operational details of a train before and after traversing a station. Its purpose is to demonstrate the feasibility of implementing the DANRT system, contingent upon the resolution of technical availability associated with MAV or RCSD technologies.

Fig. A.1 shows a schematic diagram of some possible operational details of a train before and after passing through a station in the DANRT system. Note that when calculating running time and distance, we only consider the worst-case scenario where all carriages need to be decoupled as shown in Fig. A.1. In Fig. A.1(a), when a train is about to pass through a station, the carriages arriving at their destination, which are represented by light red squares, are detached from the train at operational speed. According to the train timetable and length of Batong Line, we can obtain that the average speed of a train v_{nor} is 12 m/s. Note that the maximum speed of a train can be larger than the average speed of a train in reality. Train length is set to 120 m. The maximum acceleration of a train a is set to 1 m/s^2 (Felez et al., 2019). The maximum speed at which a train crosses tracks v_{cross} is set to 7 m/s for safe operation of the system according to the *Code for design of metro GB50157-2013* of China. All carriages will then go through four phases.

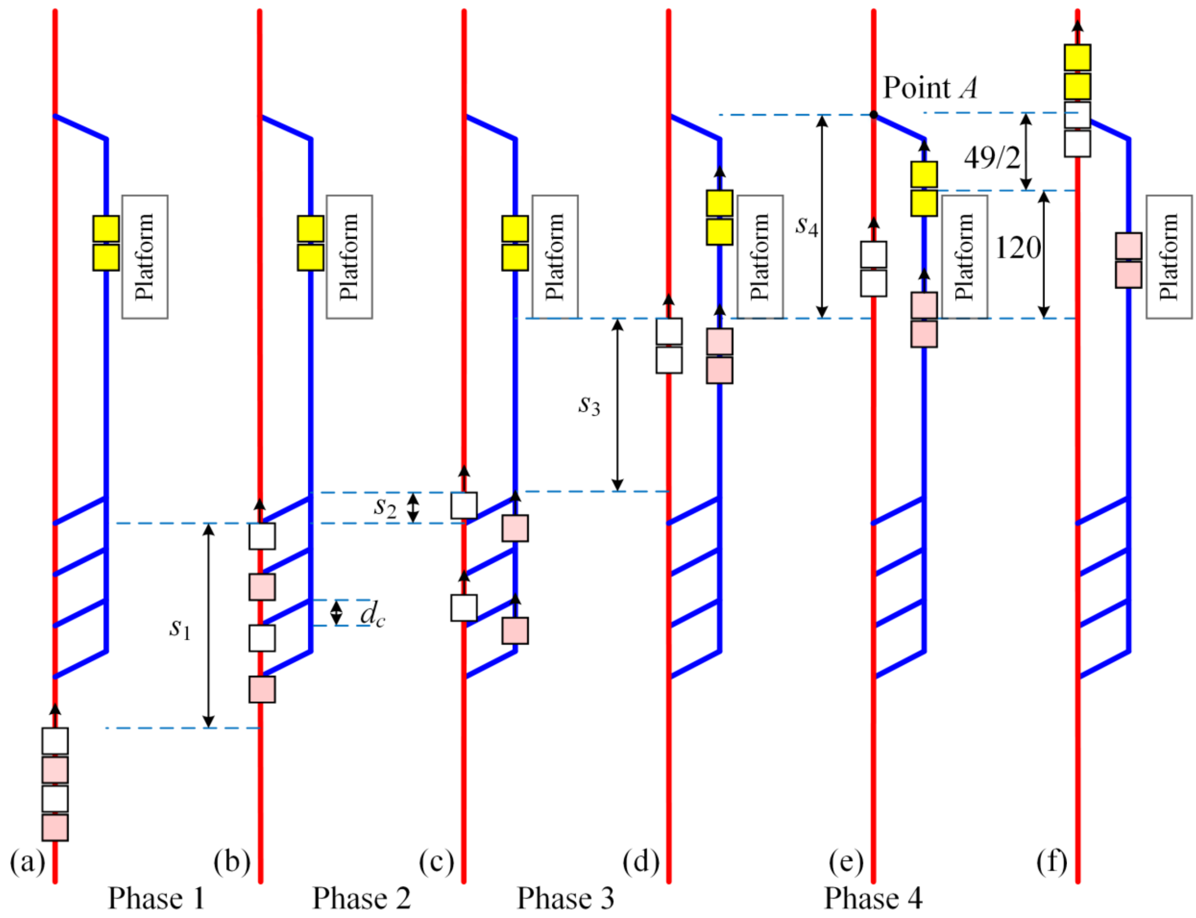


Fig. A.1. A schematic diagram of some possible operational details of a train before and after passing through a station in the DANRT system. When a train is about to pass through a station, (a) carriages arriving at their destination are detached. All carriages will go through (b) phase 1, (c) phase 2, (d) phase 3, and (e) phase 4. (f) The operation process of the train ends.

Phase 1: All carriages slow down and keep a sufficient distance until running to the track change point as shown in Fig. A.1(b). Let s_1 represent the distance traveled by the first carriage during this process, t_1 denote the time required for this process, T_{tra} be the maximum number of carriages a train can load at the same time. The speed of the train after deceleration is v_{small}. Note that the operating time of a conventional electric switch machine can be no more than 0.6 s (Qian et al., 2019). Therefore, the minimum distance between two adjacent carriages $d_c = 0.6 \cdot v_{small}$. In this process, the first carriage moves $\left(T_{\mathrm{tra}} - 1\right) \cdot 0.6 \cdot v_{small}$ further than the last carriage. As shown in Fig. A.2, the last carriage decelerates with maximum acceleration and maintains a small speed until the distance from the first carriage is $(T_{\mathrm{tra}}-1)\cdot 0.6\cdot v_{small}$. In other words, the area of the shaded part in Fig. A.2 is $(T_{\mathrm{tra}}-1)\cdot 0.6\cdot v_{small}$, i.e., $(t_1-12+v_{small})\cdot(12-v_{small})/2=(T_{\mathrm{tra}}-1)\cdot 0.6\cdot v_{small}$. We can obtain the time required for this process

$$t_1=\frac{2\cdot(T_{\mathrm{tra}}-1)\cdot 0.6\cdot v_{small}}{(12-v_{small})}+12-v_{small}.$$

The distance traveled by the first carriage during this process can also be calculated as

$$s_1=\frac{(v_{small}+12)\cdot t_1}{2}.$$

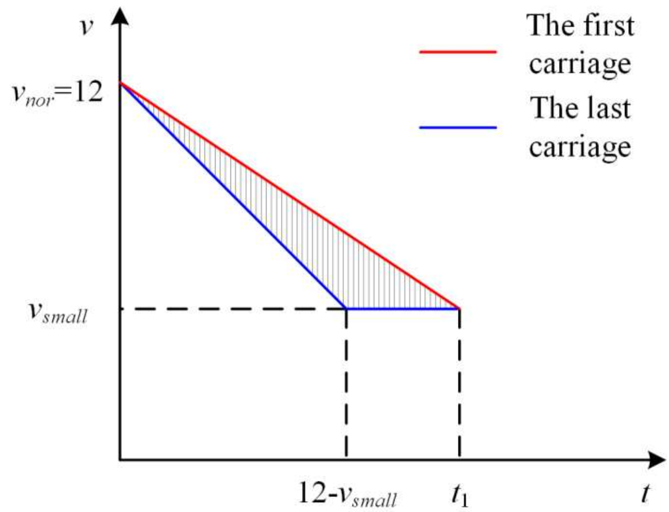


Fig. A.2. Speed-time profiles of the first and the last carriages in phase 1.

Phase 2: At the beginning of phase 2, the positions of carriages are shown in Fig. A.1(b), i.e., the head of each carriage has reached the guide tracks and the electric switch machine has been set up. During phase 2, carriages that are about to arrive at their destination shift to guide tracks, while the other carriages continue traveling on the main tracks as shown in Fig. A.1(c). For example, the 2nd carriage is about to shift to guide tracks as shown in Fig. A.1(b). If this carriage is completely transferred from the main tracks to the guide tracks, i.e., the rear of the carriage also reaches the guide tracks already, the carriage then needs to travel at least a one-carriage-length distance. Phase 2 ends when all these carriages that are about to arrive at their destination are completely transferred to the guide tracks. Because the speed of each carriage is equal, the first carriage also travels a one-carriage-length distance during this process. Note that there may be multiple guide tracks and conventional electric switch machines to allow decoupling of the carriages in the middle of the train. If a carriage in the middle of a train needs to be decoupled, the corresponding conventional electric switch machine will help the carriage shift to guide tracks at the beginning of phase 2 within 0.6 s. Let s_2 represent the distance traveled by the first carriage during this process and t_2 denote the time required for this process. The carriage length is $120/T_{\text{tra}}$, and thus we can obtain that the minimum distance traveled by the first carriage during this process s_2 is $120/T_{\text{tra}}$. The time required for this process t_2 can be calculated as $120/T_{\text{tra}}/v_{small}$.

Phase 3: Carriages on the two track lines are recoupled by accelerating the rear carriages as shown in Fig. A.1(d). We consider the worst-case scenario where only the first and the T_{tra}-th carriages need to be coupled. At the beginning of this process, the distance between the rear of the first carriage and the front of the T_{tra}-th carriage is the sum of the distance between two adjacent carriages plus the sum of the length of all the other $T_{\text{tra}}-2$ carriages, i.e., $(T_{\text{tra}}-1)\cdot 0.6\cdot v_{small}+(T_{\text{tra}}-2)\cdot 120/T_{\text{tra}}$. In other words, at the end of this process, the T_{tra}-th carriage should travel $(T_{\text{tra}}-1)\cdot 0.6\cdot v_{small}+(T_{\text{tra}}-2)\cdot 120/T_{\text{tra}}$ farther than the first carriage. Note that if all carriages need to be coupled, the distance between the rear of the first carriage and the front of the last carriage equals the area of the shaded part in Fig. A.2. As shown in Fig. A.3, when the maximum speed of the T_{tra}-th carriage cannot reach 12 m/s, the distance that the T_{tra}-th carriage travels more than the first carriage is represented by the area of the triangle (i.e., shaded part) in Fig. A.3(a). The time required for phase 3 t_3 can be calculated as

$$t_3 = 2\sqrt{(T_{\text{tra}}-1)\cdot(120/T_{\text{tra}}+0.6\cdot v_{small})-120/T_{\text{tra}}}\ .$$

Similarly, in the case shown in Fig. A.3(b), the time required for phase 3 t_3 can be calculated as

$$t_3 = \frac{(T_{\text{tra}}-1)\cdot(120/T_{\text{tra}}+0.6\cdot v_{small})-120/T_{\text{tra}}}{12-v_{small}}+12-v_{small}\,.$$

The distance traveled by the first carriage during this process can be calculated as $s_3 = v_{small}\cdot t_3$.

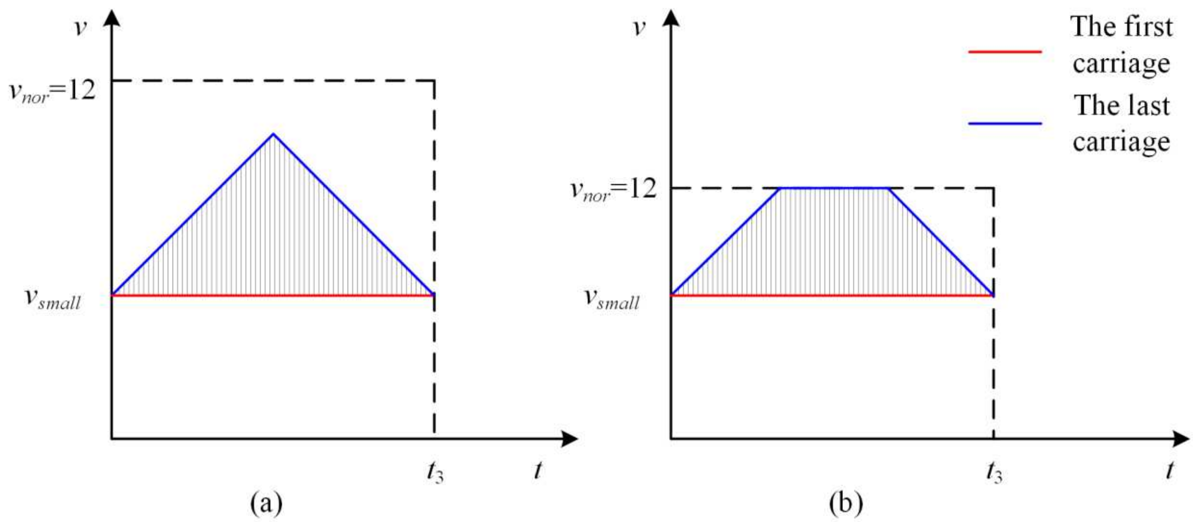


Fig. A.3. Speed-time profiles of the first and the last carriages in phase 3 when the maximum speed of the T_{tra}-th carriage (a) cannot reach 12 m/s and (b) can reach 12 m/s.

Phase 4: Carriages parked at the platform represented by yellow squares in Figs. A.1(d) and A.1(e) accelerate early and reach a speed of $v_{cross}=7$ m/s before shifting to the main track line. This means that the length of the guide track line should be 49/2 m as shown in Fig. A.1(f). When carriages on guide tracks shift to the main tracks at point *A* as shown in Fig. A.1(e), the carriages on the main tracks just arrive at point *A*. These two sets of carriages are then coupled to form a whole. Those carriages then accelerate to 12 m/s. The carriages on main tracks should arrive at point *A* as soon as possible to shorten the headway. Therefore, these carriages should change speed as shown in Fig. A.4. In the process, the carriages on the main track line travel $s_4=120+49/2$ m, which can be represented by the area of the shaded part in Fig. A.4. Similar to the calculation method in phase 3, the time required for phase 4 t_4 in Fig. A.4(a) can be calculated as

$$t_4=2\sqrt{120+{v_{small}}^2/2+{v_{cross}}^2}-v_{small}-2\cdot v_{cross}+12\,.$$

In the case shown in Fig. A.4(b), the time required for phase 3 t_4 can be calculated as

$$t_4=34-v_{small}+{v_{small}}^2/24+{v_{cross}}^2/12-2\cdot v_{cross}\,.$$

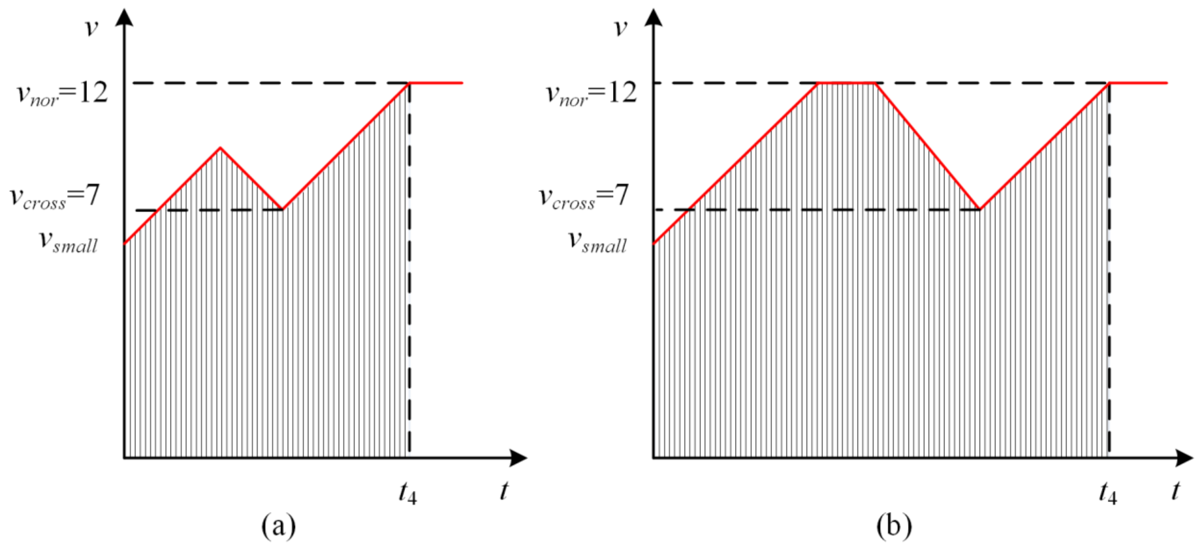


Fig. A.4. Speed-time profiles of the carriages on the main track line in phase 4 when speed (a) cannot reach 12 m/s and (b) can reach 12 m/s.

The total time required for these four phases $t_{DANRT} = t_1 + t_2 + t_3 + t_4$ and the total distance traveled by the train $s_{DANRT} = s_1 + s_2 + s_3 + s_4$. If a train in the traditional system passes through a station and travels such a distance, the required time can be calculated by $t_{tra} = t_{acc} + t_{remain} + t_{stop}$, where t_{acc} represents the time required for acceleration and deceleration when the train stops at the station, t_{remain} denotes the time for the train to travel the remaining distance at operational speed, and t_{stop} represents the duration of the train stopping at the platform. The operational speed of a train is also 12 m/s and maximum acceleration of a train is 1 m/s^2, thus $t_{acc} = 12 + 12 = 24$ s. The distance traveled by the train during this process $s_{acc} = 72 + 72 = 144$ m. If the train travels s_{DANRT}, the remaining distance $s_{remain} = s_{DANRT} - s_{acc}$ and thus $t_{remain} = s_{remain}/12$. t_{stop} is set to 38 s according to a set of video recordings taken at the Xierqi station in Beijing Subway. As a result, we can obtain that compared with the traditional system, passengers in the new system will save $t_{save} = t_{tra} - t_{DANRT}$ for each station they pass through. Table A.1 shows t_{save}-values when v_{small} and T_{tra} take different values.

Table A.1 t_{save}-values (unit: s) when v_{small} (unit: m/s) and T_{tra} take different values

v_{small}	$T_{\text{tra}} = 6$	$T_{\text{tra}} = 8$	$T_{\text{tra}} = 10$	$T_{\text{tra}} = 12$	$T_{\text{tra}} = 15$	$T_{\text{tra}} = 20$	$T_{\text{tra}} = 40$
1	-1.71	1.84	3.89	5.19	6.39	7.39	7.64
2	11.29	12.14	12.49	12.59	12.49	11.99	8.74
3	17.29	17.11	16.76	16.33	15.59	14.26	8.51

4	21.29	20.49	19.69	18.89	17.69	15.69	7.69
5	24.29	23.04	21.89	20.79	19.19	16.59	6.44
6	26.63	25.01	23.56	22.19	20.23	17.06	4.81
7	28.43	26.50	24.78	23.16	20.85	17.13	NaN

Note: NaN means that when v_{small} and T_{tra} take the corresponding values, the total distance traveled by the train s_{DANRT} exceeds 1000 m, which is assumed to be the minimum distance between two adjacent stations.

We then calculate the minimum departure interval (headway) required to prevent a train from rear-end collision on main tracks. When the current train passes through a station, the distance between it and the next train will be shortened by $s_{shorten}$, which can be represented by the area of the shaded part in Fig. A.5. The calculation method is similar to that in the four phases mentioned above. We can then obtain the minimum departure interval

$$t_{dep} = \frac{s_{shorten} + 120}{12}.$$

Table A.2 shows t_{dep}-values when v_{small} and T_{tra} take different values.

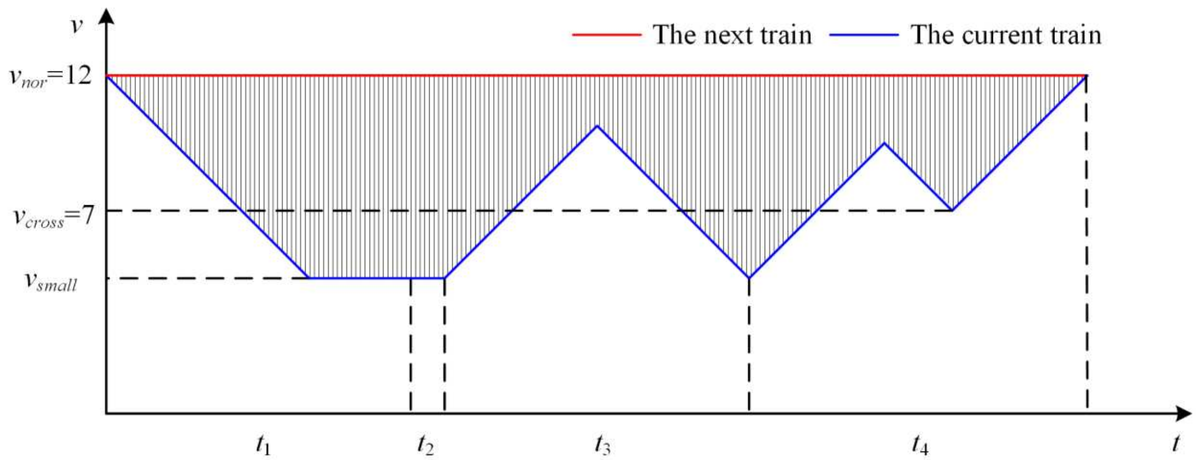


Fig. A.5. Speed-time profiles of the next train and the T_{tra}-th carriage of the current train.

Table A.2 t_{dep}-values (unit: s) when v_{small} (unit: m/s) and T_{tra} take different values

v_{small}	$T_{\text{tra}}=6$	$T_{\text{tra}}=8$	$T_{\text{tra}}=10$	$T_{\text{tra}}=12$	$T_{\text{tra}}=15$	$T_{\text{tra}}=20$	$T_{\text{tra}}=40$
1	51.08	46.70	44.15	42.52	40.98	39.65	38.90
2	38.08	36.40	35.55	35.12	34.88	35.05	37.80
3	32.08	31.43	31.28	31.38	31.78	32.78	38.03
4	28.08	28.05	28.35	28.82	29.68	31.35	38.85

5	25.08	25.50	26.15	26.92	28.18	30.45	40.10
6	22.75	23.53	24.48	25.52	27.15	29.98	41.73
7	20.94	22.04	23.26	24.55	26.53	29.91	NaN

Note: NaN means that when v_{small} and T_{tra} take the corresponding values, the total distance traveled by the train s_{DANRT} exceeds 1000 m, which is assumed to be the minimum distance between two adjacent stations.

Finally, we calculate the minimum departure interval (headway) required to prevent a train from rear-end collision on guide tracks. Carriages detached from a train in phase 4 need to change speed as shown in Fig. A.6. When the 3rd train passes through the station, the carriages that are about to be attached to the train accelerate to a speed that can be coupled with the train as shown in Fig. A.6. The distance between carriages detached from the current train on guide tracks and the carriages detached from the 3rd train will be shortened by $s_{shorten}^{guide}$, which can be represented by the area of the shaded part in Fig. A.6. The calculation method is similar to that in the four phases mentioned above. We can then obtain the minimum departure interval

$$t_{dep}^{guide} = \frac{s_{shorten}^{guide} + 120}{v_{small}}.$$

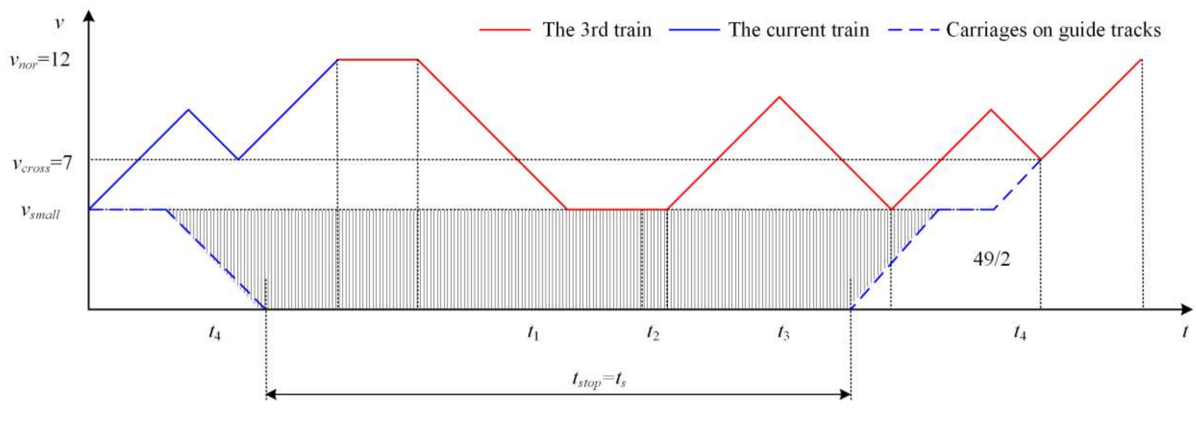


Fig. A.6. Speed-time profiles of the carriages on guide tracks.

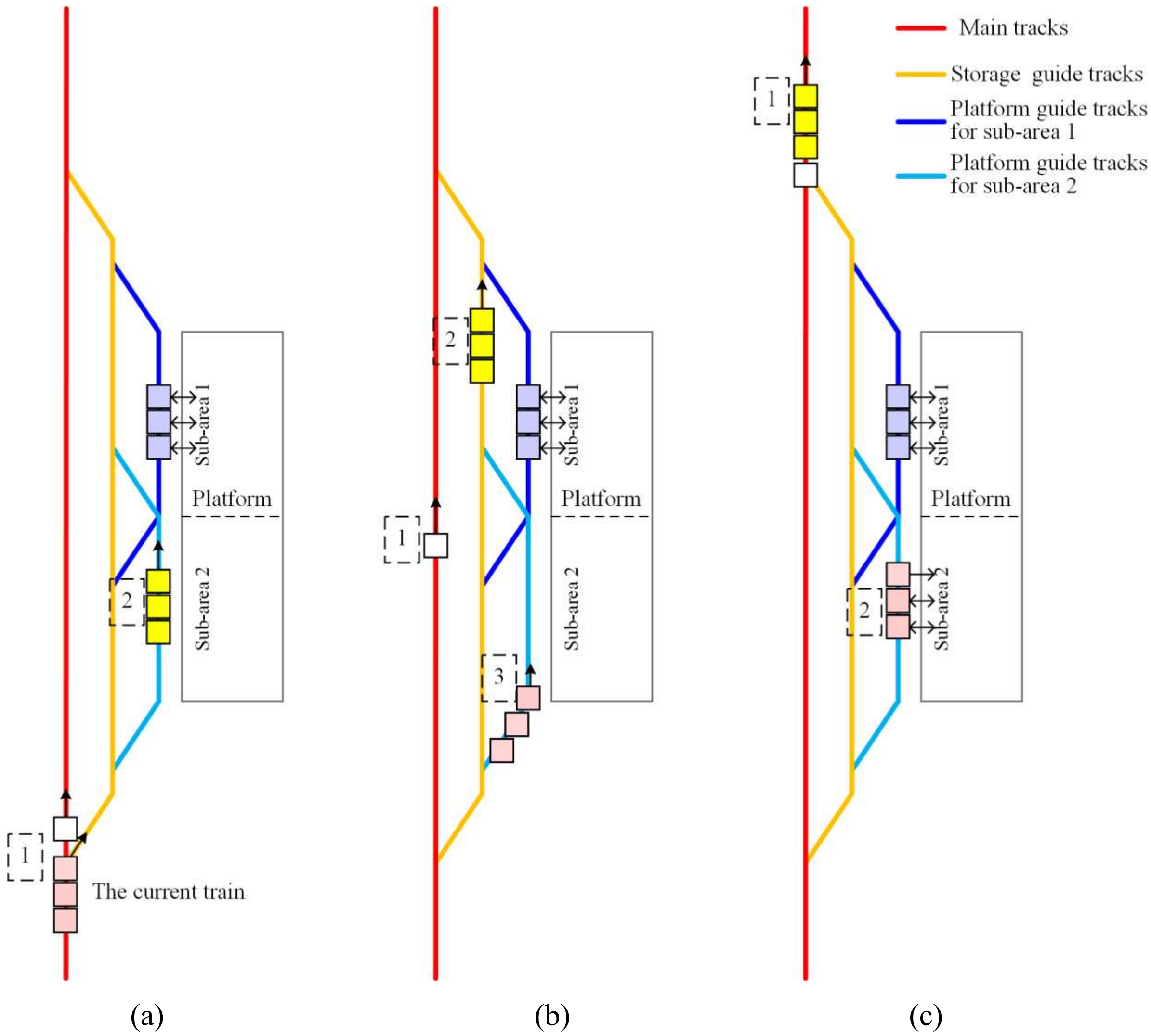
We set $v_{small} = 6$, t_{dep}^{guide} can be calculated by $(38 \times 6 + 6 \times 6 + 120)/6 = 64$. In other words, the departure interval between the 1st and 3rd trains is 64 s. It is worth mentioning that, when the carriages detached from the 2nd train travel on the guide tracks, the carriages detached from the 1st train have been parked at the corresponding sub-area without affecting the operation of the carriages detached from the 2nd train. Thus, the headway between two successive trains should be at least $64/2=32$ s.

Appendix B: Carriage operation process in the double-area system

This appendix provides a more detailed operation process of carriages passing through a station in the double-area system.

Fig. B.1 shows the carriage operation process of two consecutive trains passing through a station in the double-area system. Storage lines are not shown in Fig. B.1. Without loss of generality, we assume that the carriages arriving at their destination in the current train stop at sub-area 2. When the current train passes through the station (see position 1 in Fig. B.1(a)), carriages arriving at their destination, which are represented by light red squares, are detached from the train. The carriages parked at sub-area 2 represented by yellow squares are ready to run (see position 2 in Fig. B.1(a)). The carriages of the current train then go through phases 1, 2, and 3 (see Appendix A). In phase 4, those carriages arriving at their destination are parked at sub-area 2 (see position 3 in Fig. B.1(b) and position 2 in Fig. B.1(c)). Meanwhile, the current train continues to run on the main tracks (see position 1 in Fig. B.1(b)). The carriages parked at sub-area 2 run along the platform guide tracks for sub-area 2 in advance (see position 2 in Fig. B.1(b)). Subsequently, those carriages that need to be attached to the current train shift to the main tracks at operational speed. Note that those carriages are in the front of the current train. Those carriages are then attached to the current train to form a whole as shown in Fig. B.1(c).

Similarly, when the next train passes through the station (see position 1 in Fig. B.1(d)), carriages arriving at their destination, which are represented by light green squares, are also detached from the train. The carriages parked at sub-area 1 represented by light purple squares are ready to run (see position 2 in Fig. B.1(d)). If there is an extra vacant carriage that needs to be attached to the train (see position 3 in Fig. B.1(d)), the carriage begins to run in advance to couple with the carriages represented by light purple squares on guide tracks (see position 1 in Fig. B.1(e)). These coupled carriages on guide tracks are attached to the train to form a whole (see position 1 in Fig. B.1(f)). Meanwhile, the carriages arriving at their destination are parked at sub-area 1 (see position 2 in Fig. B.1(e) and position 2 in Fig. B.1(f)).


Main tracks
Storage guide tracks
Platform guide tracks for sub-area 1
Platform guide tracks for sub-area 2
The current train
Platform
Sub-area 1
Sub-area 2
1
2
3
(a)
(b)
(c)

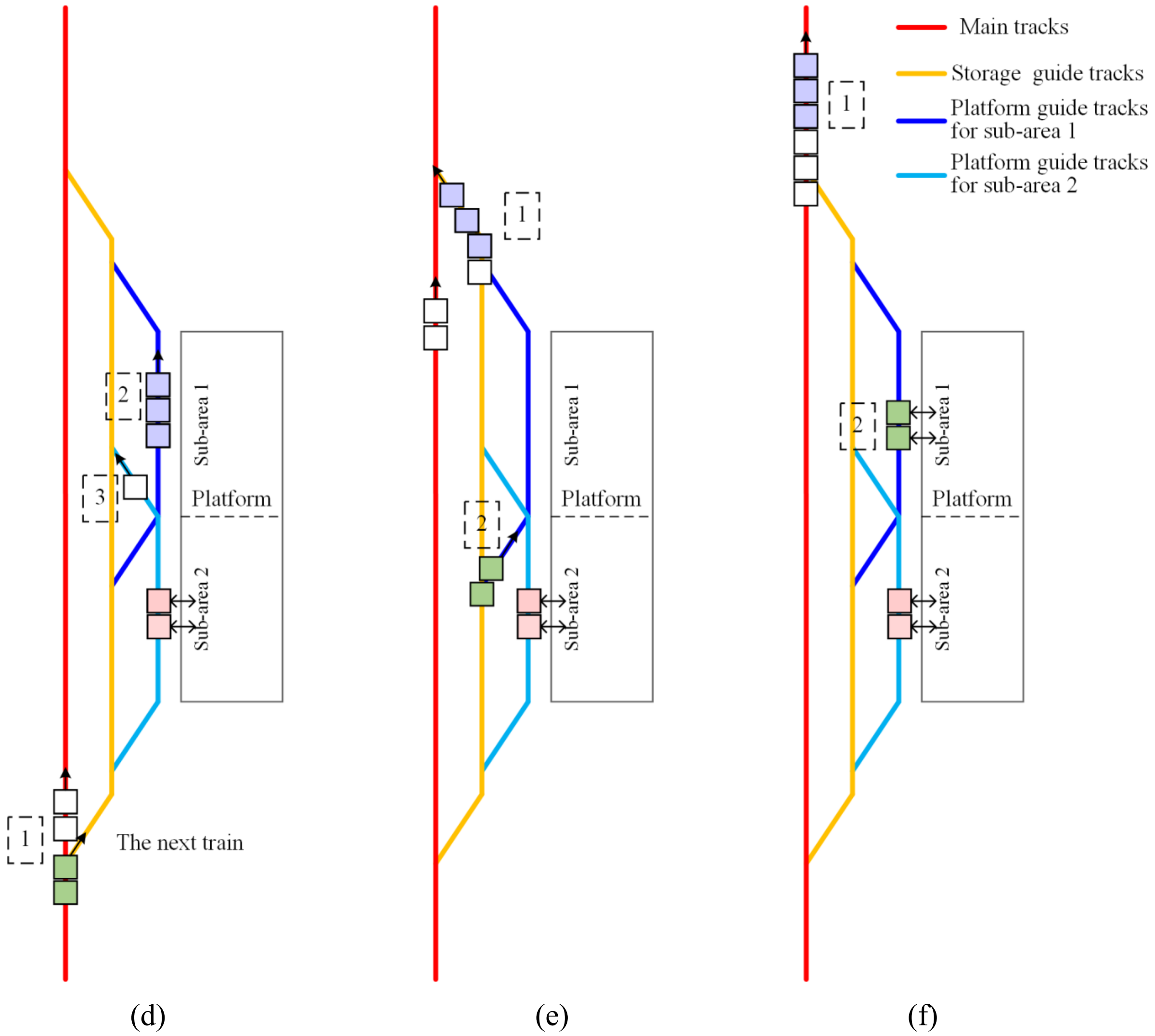


Fig. B.1. Carriage operation process of two consecutive trains passing through a station in the double-area system. (a) The current train is about to pass through the station. (b) The current train is passing through the station. (c) The current train leaves the station. (d) The next train is about to pass through the station. (e) The next train is passing through the station. (f) The next train leaves the station.

References

Ansari Esfeh, M., Wirasinghe, S. C., Saidi, S., Kattan, L., 2021. Waiting time and headway modelling for urban transit systems – a critical review and proposed approach. Transport Reviews 41 (2), 141-163.

Barrena, E., Canca, D., Coelho, L.C., Laporte, G., 2014. Single-line rail rapid transit timetabling under dynamic passenger demand. Transportation Research Part B: Methodological 70, 134-150.

Behiri, W., Belmokhtar-Berraf, S., Chu, C., 2018. Urban freight transport using passenger rail network: Scientific issues and quantitative analysis. Transportation Research Part E: Logistics and Transportation Review 115, 227-245.

Bhoopalam, A.K., Agatz, N., Zuidwijk, R., 2018. Planning of truck platoons: A literature review and directions for future research. Transportation Research Part B: Methodological 107, 212-228.

Boysen, N., Briskorn, D., Schwerdfeger, S., 2018. The identical-path truck platooning problem. Transportation Research Part B: Methodological 109, 26-39.

Ceder, A., 2021. Urban mobility and public transport: future perspectives and review. International Journal of Urban Sciences 25 (4), 455-479.

Chen, Z., Li, X., Qu, X., 2022. A continuous model for designing corridor systems with modular autonomous vehicles enabling station-wise docking. Transportation Science 56 (1), 1-30.

Chen, Z. and Li, X., 2021. Designing corridor systems with modular autonomous vehicles enabling station-wise docking: Discrete modeling method. Transportation Research Part E: Logistics and Transportation Review 152, 102388.

Chen, Z., Li, X., Zhou, X., 2020. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Continuous modeling method. Transportation Research Part B: Methodological 132, 76-100.

Daganzo, C.F., 2022. An operating system for extra long urban trains. Transportation Research Part B: Methodological 158, 323-340.

Dai, Z., Liu, X.C., Chen, X., Ma, X., 2020. Joint optimization of scheduling and capacity for mixed traffic with autonomous and human-driven buses: A dynamic programming approach. Transportation Research Part C: Emerging Technologies 114, 598-619.

Farhat, N., Ward, C.P., Dixon, R., Goodall, R.M., 2018. Benefits of mechatronically guided vehicles on railway track switches. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 234 (3), 276-288.

Felez, J., Kim, Y., Borrelli, F., 2019. A model predictive control approach for virtual coupling in railways. IEEE Transactions on Intelligent Transportation Systems 20 (7), 2728-2739.

Goodall, R.M., 2011. Control engineering challenges for railway trains of the future, Measurement and Control 44, 16-24.

Guo, Q.W., Chow, J.Y.J., Schonfeld, P., 2017. Stochastic dynamic switching in fixed and flexible transit services as market entry-exit real options. Transportation Research Procedia 23, 380-399.

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. Applied Mathematical Modelling 40(17), 8048-8066.

Han, Y., Zhang, T., Wang, M., 2020. Holiday travel behavior analysis and empirical study with integrated travel reservation information usage. Transportation Research Part A: Policy and Practice 134, 130-151.

Jamili, A. and Pourseyed Aghaee, M., 2015. Robust stop-skipping patterns in urban railway operations under traffic alteration situation. Transportation Research Part C: Emerging Technologies 61, 63-74.

Jiang, Z., Fan, W., Liu, W., Zhu, B., Gu, J., 2018. Reinforcement learning approach for coordinated passenger inflow control of urban rail transit in peak hours. Transportation Research Part C: Emerging Technologies 88, 1-16.

Kamel, I., Shalaby, A., Abdulhai, B., 2020. A modelling platform for optimizing time-dependent transit fares in large-scale multimodal networks. Transport Policy 92, 38-54.

Li, S., Dessouky, M.M., Yang, L., Gao, Z., 2017. Joint optimal train regulation and passenger flow control strategy for high-frequency metro lines. Transportation Research Part B: Methodological 99, 113-137.

Liu, S., Yao, E., Li, B., 2019. Exploring urban rail transit station-level ridership growth with network expansion. Transportation Research Part D: Transport and Environment 73, 391-402.

Liu, Z., Correia, G.H. de A., Ma, Z., Li, S., Ma, X., 2023. Integrated optimization of timetable, bus formation, and vehicle scheduling in autonomous modular public transport systems. Transportation Research Part C: Emerging Technologies 155, 104306.

Nold, M., 2019. Kupplungsunterstützungsvorrichtung zum An- und Abkuppeln von Schienenfahrzeugen, in: Deutsches Patent- und Markenamt, DE102018009589B3.

Nold, M. and Corman, F., 2021. Dynamic train unit coupling and decoupling at cruising speed: Systematic classification, operational potentials, and research agenda. Journal of Rail Transport Planning & Management 18, 100241.

Qian, K., Li, J., Luo, M., 2019. Current situation and development of comprehensive automation technology in marshalling station. IOP Conference Series: Materials Science and Engineering 688 (4), 044028.

Sala, M. and Soriguera, F., 2021. Capacity of a freeway lane with platoons of autonomous vehicles mixed with regular traffic. Transportation Research Part B: Methodological 147, 116-131.

Schweizer, J., Mantecchini, L., Greenwood, J., 2011. Analytical capacity limits of personal rapid transit stations. Automated People Movers and Transit Systems 2011: From People Movers to Fully Automated Urban Mass Transit. American Society of Civil Engineers (ASCE), 326-338.

Schwerdfeger, S., Otto, A., Boysen, N., 2021. Rail platooning: Scheduling trains along a rail corridor with rapid-shunting facilities. European Journal of Operational Research 294 (2), 760-778.

Shafahi, Y. and Khani, A., 2010. A practical model for transfer optimization in a transit network: Model formulations and solutions. Transportation Research Part A: Policy and Practice 44 (6), 377-389.

Shi, J., Qin, T., Yang, L., Xiao, X., Guo, J., Shen, Y., Zhou, H., 2022. Flexible train capacity allocation for an overcrowded metro line: A new passenger flow control approach. Transportation Research Part C: Emerging Technologies 140, 103676.

Shi, J., Yang, L., Yang, J., Zhou, F., Gao, Z., 2019. Cooperative passenger flow control in an oversaturated metro network with operational risk thresholds. Transportation Research Part C: Emerging Technologies 107, 301-336.

Shi, J., Yang, J., Yang, L., Tao, L., Qiang, S., Di, Z., Guo, J., 2023. Safety-oriented train timetabling and stop planning with time-varying and elastic demand on overcrowded commuter metro lines. Transportation Research Part E: Logistics and Transportation Review 175, 103136.

Shi, X. and Li, X., 2021. Operations design of modular vehicles on an oversaturated corridor with first-in, first-out passenger queueing. Transportation Science 55 (5), 1187-1205.

Sun, L., Jin, J.G., Lee, D.H., Axhausen, K.W., Erath, A., 2014. Demand-driven timetable design for metro services. Transportation Research Part C: Emerging Technologies 46, 284-299.

Tian, Q., Lin, Y.H., Wang, D.Z.W., Liu, Y., 2022. Planning for modular-vehicle transit service system: Model formulation and solution methods. Transportation Research Part C: Emerging Technologies 138, 103627.

Tian, Q., Yang, H., Huang, H.J., 2012. Pareto efficient strategies for regulating public transit operations. Public Transport 3 (3), 199-212.

United Nations, 2019. World Urbanization Prospects: The 2018 Revision.

Visser, J.G.S.N., 2018. The development of underground freight transport: An overview.

Tunnelling and Underground Space Technology 80, 123-127.
Xu, X.Y., Liu, J., Li, H.Y., Jiang, M., 2016. Capacity-oriented passenger flow control under uncertain demand: Algorithm development and real-world case study. Transportation Research Part E: Logistics and Transportation Review 87, 130-148.
Yang, H. and Tang, Y., 2018. Managing rail transit peak-hour congestion with a fare-reward scheme. Transportation Research Part B: Methodological 110, 122-136.
Yang, X., Xue, Q., Yang, X., Yin, H., Qu, Y., Li, X., Wu, J., 2021. A novel prediction model for the inbound passenger flow of urban rail transit. Information Sciences 566, 347-363.
Yin, J., Tang, T., Yang, L., Gao, Z., Ran, B., 2016. Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: An approximate dynamic programming approach. Transportation Research Part B: Methodological 91, 178-210.
Yuan, F., Sun, H., Kang, L., Lv, Y., Yang, X., Wei, Y., 2023. An integrated optimization approach for passenger flow control strategy and metro train scheduling considering skip-stop patterns in special situations. Applied Mathematical Modelling 118, 412-436.
Yuan, F., Sun, H., Kang, L., Wu, J., 2020. Passenger flow control strategies for urban rail transit networks. Applied Mathematical Modelling 82, 168-188.
Yuan, Y., Li, S., Yang, L., Gao, Z., 2022. Real-time optimization of train regulation and passenger flow control for urban rail transit network under frequent disturbances. Transportation Research Part E: Logistics and Transportation Review 168, 102942.
Zhang, T., Li, D., Qiao, Y., 2018. Comprehensive optimization of urban rail transit timetable by minimizing total travel times under time-dependent passenger demand and congested conditions. Applied Mathematical Modelling 58, 421-446.
Zhang, Z., Tafreshian, A., Masoud, N., 2020. Modular transit: Using autonomy and modularity to improve performance in public transportation. Transportation Research Part E: Logistics and Transportation Review 141, 102033.
Zhao, L., Zhou, J., Li, H., Yang, P., Zhou, L., 2021. Optimizing the design of an intra-city metro logistics system based on a hub-and-spoke network model. Tunnelling and Underground Space Technology 116, 104086.