Int. J. Appl. Math. Comput. Sci., 2022, Vol. 32, No. 4, 553–567
DOI: 10.34768/amcs-2022-0039

A HYBRID CONTROL STRATEGY FOR A DYNAMIC SCHEDULING PROBLEM IN TRANSIT NETWORKS

ZHONGSHAN LIU [a], BIN YU [a], LI ZHANG [a], WENSI WANG [a,*]

[a] School of Transportation Science and Engineering
Beihang University
Beijing 100191, PR China
e-mail: wensiwws@buaa.edu.cn

Public transportation is often disrupted by disturbances, such as the uncertain travel time caused by road congestion. Therefore, the operators need to take real-time measures to guarantee the service reliability of transit networks. In this paper, we investigate a dynamic scheduling problem in a transit network, which takes account of the impact of disturbances on bus services. The objective is to minimize the total travel time of passengers in the transit network. A two-layer control method is developed to solve the proposed problem based on a hybrid control strategy. Specifically, relying on conventional strategies (e.g., holding, stop-skipping), the hybrid control strategy makes full use of the idle standby buses at the depot. Standby buses can be dispatched to bus fleets to provide temporary or regular services. Besides, deep reinforcement learning (DRL) is adopted to solve the problem of continuous decision-making. A long short-term memory (LSTM) method is added to the DRL framework to predict the passenger demand in the future, which enables the current decision to adapt to disturbances. The numerical results indicate that the hybrid control strategy can reduce the average headway of the bus fleet and improve the reliability of bus service.

Keywords: service reliability, transit network, proactive control method, deep reinforcement learning, hybrid control strategy.

1. Introduction

1.1. Background. Public transportation is often disrupted by disturbances (e.g., the uncertain travel time caused by road congestion). These disturbances usually lead to bus bunching and large intervals, which would result in prolonged travel time and inconvenience. To guarantee the attractiveness of bus services, bus operators strive to promote the regularity and punctuality of bus services. Therefore, it is crucial for the bus operators to reduce the impact of disturbances on bus services.

The conventional control strategies (e.g., holding, stop-skipping) are mainly applied to a single bus line. Application of these strategies to a transit network is complex. Due to the limited amount of buses, large passenger demand and long driving time in a single trip, the level of bus service of some lines may be poor. This problem would become prominent during rush hours. If the bus services cannot meet the passenger demand based on conventional control strategies, another strategy need to be considered to improve the service reliability of bus lines. How to design a reasonable and novel strategy to solve the dynamic scheduling problems is attractive to many researchers (Petit *et al.*, 2019; Argote-Cabanero *et al.*, 2015; Wang *et al.*, 2020).

The existing literature may not fully consider the characteristics of standby buses. In practice, some idle buses are placed at the depot by the bus company. These standby buses could be flexibly inserted into the bus fleet and provide services for passengers. After arriving at the terminal stop and ending a single service, a standby bus would return to the standby location and wait for the next task. If the service quality of multiple bus lines decreases due to the disturbances, there may not be enough standby buses to serve passengers on all bus lines. It is necessary to formulate an optimal scheme to maximize the benefits brought by standby buses. Using the standby buses to provide services can not only make use of idle resources, but also help to solve the scheduling problem in the transit network. Compared with the service provided by standby

*Corresponding author

buses from the depot, it would reduce the scheduling times if these buses are located in initial positions where service reliability is likely poor in the future (Petit *et al.*, 2019). Besides, it may be economical to allow the operators to dynamically adjust the service mode (e.g. regular service, temporary service) of standby buses. However, in the literature communications on standby bus strategies to solve scheduling problems in transit networks are limited. The effect of a hybrid application of a standby bus strategy and a conventional control strategy to improve the quality of public transportation service is unknown. To complement this field, this paper develops hybrid control strategies to solve the dynamic scheduling problem in the transit network.

1.2. Literature review. One of the most well-known control strategies is holding. Holding is a scheduling method to delay the departure time of some buses purposefully at stops. Holding mainly contains two kinds of methods: schedule-based holding, headway-based holding. Schedule-based holding and attempts to adjust the departure time of buses according to the timetable (Zhu *et al.*, 2017). Van Oort *et al.* (2012) analyzed the bus service in the Hague, the Netherlands. The analysis shows that the schedule-based holding can save 60% of the waiting time at stops and on board for passengers. Xuan *et al.* (2011) introduced the concept of a virtual timetable. The authors determined the holding time based on the deviation between the actual arrival time of buses and the virtual timetable. Hall *et al.* (2001) discussed the problem of holding time at transfer stops with the goal of minimizing the waiting time of passengers.

For bus services with a high frequency, short average headway and high passenger demand, it is difficult for buses to provide services according to a fixed timetable. Headway-based holding is suitable for this situation. The objectives of the headway-based holding strategy can be roughly divided into two categories. One of the optimization models is established with the goal of minimizing the mean or variance of headway (Fu and Yang, 2002). The other is to minimize the total waiting or travel time of passengers (Barnett, 1974). The decision variable of the headway-based holding is the dwell time of buses at different stops. Fu and Yang (2002) took the value of headway as an important index to evaluate the performance of the bus system. The research showed that the selection of control stops would affect the performance of the system. Barnett (1974) designed a holding control model to reduce the risk of disturbances with the goal of minimizing the waiting time for passengers. They made an empirical analysis based on the actual operation data in Boston, USA.

Stop skipping is also a common control strategy. It allows some buses to skip partial stops, so as to speed up the turnover rate of buses. Stop-skipping mainly contains three kinds of methods: limited-stop (Vuchic, 1973; Larrain *et al.*, 2010), short-turning (Jordan and Turnquist, 1979; Furth, 1986; Ceder, 1988), deadheading (Cortés *et al.*, 2011; Furth, 1985; Ceder and Stern, 1981). Vuchic (1973) evaluated the limited-stop mode and found that this service can improve the driving speed of buses. Considering the influencing factors such as climate, the direction of lines, and different periods, Larrain *et al.* (2010) believed that the attraction of the limited-stop mode was affected by the distribution of passengers. The results showed that the average driving distance had a significant impact on the quality of services. The short-turning mode considers the use of buses to provide services in areas with high passenger demands. Compared with the full service, the services provided by buses only in certain road segments have shorter operation time and higher service frequency (Jordan and Turnquist, 1979; Furth, 1986; Ceder, 1988). The deadheading mode refers to the setting, where, if the service in the high demand direction is completed, the bus quickly drives to the origin stop with the form of empty bus. The bus would provide service from the origin stop again. There is relatively little research on deadheading because of the long driving distance with empty buses, which increases the operating cost (Cortés *et al.*, 2011; Furth, 1985; Ceder and Stern, 1981).

Bus substitution is a novel method. This service refers to inserting standby buses into the bus fleet to replace the early or delayed buses. Then, a retired bus begins to operate a drop-off mode to deliver the passengers on board, but passengers waiting at stops are not allowed to board (Petit *et al.*, 2018; 2019). Petit *et al.* (2018) first studied how to apply the bus substitution strategy to one bus line. This paper determined the amount and dispatch time of standby buses with the goal of minimizing the total cost of passengers and bus operators. They applied an adaptive dynamic programming method to solve the proposed problem. Then, Petit *et al.* (2019) discussed feasibility of using standby buses to serve passengers on multiple independent lines.

There are few studies on interstop control methods, which mainly focus on controlling the driving speed of buses. Daganzo and Pilachowski (2011) proposed a dynamic adaptive control model for coordinating and synchronizing the speed of bus fleets. This method can optimize the headway of bus fleets as well as the average driving speed of buses. Estrada *et al.* (2016) combined signal priority with cruising speed control. If the headway is larger than the set threshold, it may implement the signal priority strategy for the bus to reduce the interval; On the contrary, the driving speed of the bus would be reduced.

At present, few studies consider how to implement a scheduling decision in multiple lines. Argote-Cabanero *et al.* (2015) combined the holding and cruising speed control under the background of multiple lines. The study assumed that the bus fleet may be affected by external interference factors such as cars. The interaction between buses on different lines was ignored. Results showed that the proposed real-time scheduling method could effectively improve the performance of the bus system. Laskaris *et al*. (2019) proposed a control model for the multiple lines to minimize the cost of passengers. The authors applied a simulation network to test the proposed model.

Reinforcement learning (RL) is an important branch of machine learning. RL could learn from the existing experience or mistakes. This method could determine a series of actions through feedback, which is conducive to increasing the probability of achieving the goal. There are many extended studies on RL. For the specific research and expansion of RL, the reader is referred to the articles by Guan *et al*. (2021) and Farazi *et al*. (2021). By combining a neural network and RL, the performance of deep reinforcement learning becomes excellent (Li *et al.*, 2016; Alesiani *et al.*, 2018; 2017; Wang and Sun, 2020). Mnih *et al*. (2015) applied the traditional Q-learning algorithm to a deep neural network and proposed the Deep Q-Network (DQN) algorithm. The algorithm consists of multiple convolutional neural networks, and the random gradient descent method is used to optimize the network parameters (Krizhevsky *et al*., 2017; Wang and Chan, 2017; Hu *et al*., 2020). A double DQN is a further extension on the basis of the DQN. This algorithm use a double Q network to solve the overestimation problem (Wang and Tang, 2021; Luo, 2020). In the field of scheduling problems of transit networks, it is a feasible research direction to solve real-time problems based on the DDQN.

To the best of our knowledge, there is a lack of research on the dynamic scheduling problem in transit networks by combining the standby bus strategy and conventional strategies. To fill this gap, a novel DDQN algorithm and a hybrid control strategy are applied in this paper to solve real-time scheduling problems.

1.3. Contributions. In this paper, we focus on the scheduling problem in a transit network, which takes account of the impact of disturbances (e.g., uncertain travel time) on the service level of public transportation. The main contributions are threefold. First, we design a hybrid control strategy by incorporating conventional strategies and the standby bus strategy. The hybrid control strategy is a novel method to deal with the dynamic scheduling problems in large transit networks. Second, a two-layer control method is introduced in this paper to determine the application time of each strategy. We establish multiple modules to explicitly describe the state of transit networks, which is conducive to designing an optimal allocation scheme of standby buses in each period. Third, after realizing the high computing real-time requirements of solving the dynamic scheduling problems, a tailored DDQN algorithm is proposed to develop an effective control policy in the transit network. The long short-term memory (LSTM) method is considered in the DDQN algorithm to predict the service reliability of bus lines in next several time steps, which ensures the current decision may better guarantee the quality of bus services in the future.

The remainder of this paper is as follows: In Section 2, a detailed introduction to the underlying scheduling strategy and some assumptions are presented. Section 3 introduces the methodology used in this paper. Results and a discussion about the relevant results are included in Section 4. The conclusions are summarized in Section 5.

2. Problem description

In this paper, we study a realistic scheduling problem in a transit network. Five actions (i.e., scheduling strategy) are used in this paper:

1. *Skipping the control stop.* In this paper, the control stop is defined as the stop is selected to execute the control strategy according to some rules, such as the magnitude of passenger demand, whether there is enough space in the platform for buses to hold. When a bus is about to arrive at the control stop, the scheduling model is activated to optimize the scheduling actions;

2. *Skipping the next stop.* This action indicates that the bus is ordered to skip the next stop after the control stop. Passengers whose destination is the next stop need to alight at the control stop;

3. *Boarding limit.* This action means that the bus needs to depart from the control stop immediately after the alighting process is completed. Generally speaking, the boarding process takes more time than the alighting process, which is especially serious in crowded stops;

4. *Holding.* This action represents that the bus needs to stay at the control stop for a period of time. This strategy could ensure the headway of the bus fleet and improve the service level of buses. The disadvantage of this action is that it would increase the travel time of passengers on board;

5. *Adding standby buses into the bus fleet.* The decision of this action consists of two parts: one is to decide when the standby buses would provide services. The other is to decide the number of standby buses to provide services at the control stop. Some basic symbols are summarized in Table 1.

Table 1. Some basic notation.

Indices and sets	Definition
S	set of standby buses
B	set of regular buses
V	set of buses in the transit network, $V = S \cup B$
N	set of stops in the transit network
R	set of routes in the transit network
k	index of a period
P	set of days in one week
W	set of weeks
K	set of periods
Parameters	**Definition**
C	capacity of buses
N_r	total number of stops of the route r
M	total number of regular buses
M_s	total number of standby buses
M_r	number of regular buses serving route r

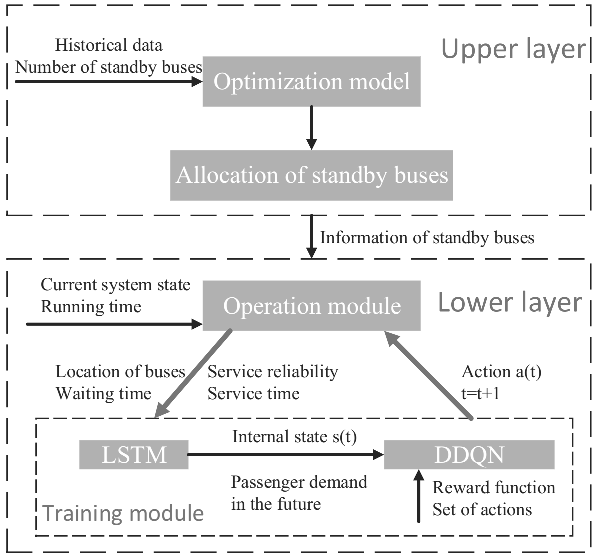


Fig. 1. Framework of the two-layer control method.

Some assumptions made in this paper are as follows:

A1. We assume that all buses in the transit network are not allowed to overtake.

A2. At each bus stop, the boarding and alighting processes are carried out at the same time. The time of boarding and alighting follows a linear formula related to the number of boarding and alighting passengers. The service time of a bus at a stop is the greater of the boarding and alighting times.

A3. All buses in route r follow a sequential order, where bus $i+1$ follows bus i, $i = 1, 2, \ldots, M_r - 1$, and bus 1 follows bus M_r.

A4. The action of skipping the control stop or the next stop is updated in real time. It might be difficult to prevent passengers whose destination is the control stop from boarding in advance. Therefore, we assume that any passenger could board at the upstream stop of the skipped stop.

A5. It is acceptable that passengers could not alight at the destination due to the stop-skipping action. The extra travel time of passengers who could not alight would be given a high penalty.

The assumptions in this paper characterize the elements of the transit network including buses, passengers, and stops.

3. Two-layer control method

The number of standby buses at different standby locations is calculated in the upper layer. The results from the upper layer are transferred to the lower layer. Then, the best action is selected in the lower layer based on the combination of DRL and LSTM. The objective of this paper is to minimize the total travel time of passengers, which has a great impact on the travel choice of passengers (Cao *et al.*, 2022). The framework of the two-layer control method is shown in Fig. 1.

3.1. Formulation of the upper layer. The core concept of the upper layer is to determine the amount of standby buses at different standby locations. The symbols used in the upper layer are shown in Table 2.

In this part, $\theta_l^{w,p}$ and $\zeta_r^{w,p}$ are determined based on historical data. It is worth noticing that the larger the value of $\zeta_r^{w,p}$, the worse the service reliability of bus lines (Fang *et al.*, 2022). More standby buses may be allocated to bus lines with poor service reliability to provide temporary services. How to design an optimal solution under the condition of limited amount of standby buses to obtain the maximum benefit is crucial. The model in the upper layer is shown as follows.

$$\max \sum_{w\in W}\sum_{r\in R}\sum_{l\in L} \eta_l^{w,p}\cdot\zeta_r^{w,p}\cdot\epsilon_l^r, \quad \forall p\in P. \tag{1}$$

Objective function. The objective function (1) is to improve the service reliability of bus lines in the transit network. The number of standby buses at different standby locations is optimized, subject to the following constraints:

$$\sum_{i\in S}\sum_{l\in L} X_{i,l}^{w,p} \le M_s, \quad \forall p\in P, w\in W, \tag{2}$$

Table 2. Notation in the upper layer.

Parameters	Definition
ϵ_l^r	binary, equals one if standby buses at location l can serve passengers on the bus line r; otherwise, equals zero
$\theta_l^{w,p}$	number of standby buses required for standby location l in day p of week w
$\zeta_r^{w,p}$	average service reliability of the bus line r in day p of week w
Variables	**Definition**
$X_{i,l}^{w,p}$	binary, equals one if standby bus i is assigned to standby location l in day p of week w; otherwise, equals zero
$\eta_l^{w,p}$	binary, equals one if the number of standby buses is equal to the required number of buses in day p of week w; otherwise, equals zero

$$\sum_{i\in S} X_{i,l}^{w,p} \le \alpha \cdot M_s, \quad \forall l \in L, p \in P, w \in W, \tag{3}$$

$$G \cdot (1-\eta_l^{w,p}) \ge \theta_l^{w,p} - \sum_{i\in S} X_{i,l}^{w,p}, \quad \forall p \in P, w \in W, l \in L, \tag{4}$$

$$X_{i,l}^{w,p} \in \{0,1\}, \quad \forall\, i \in S, l \in L, p \in P, w \in W, \tag{5}$$

$$\eta_l^{w,p} \in \{0,1\}, \quad \forall\, l \in L, w \in W, p \in P. \tag{6}$$

Service reliability constraint. Constraints (2) ensure that the amount of standby buses at all locations does not exceed the total amount of standby buses. Constraints (3) determine the maximum number of standby buses that can be placed in any standby locations. Here α is a parameter, which needs to be set according to the actual operation status. Constraints (4) link the variables $\eta_l^{w,p}$, $\theta_l^{w,p}$ and $x_{i,l}^{w,p}$. G is a very large (infinite) positive number. Constraints (5) and (6) describe the domains of decision variables.

The calculation results are transmitted to the lower layer. After one week, the value of $\theta_l^{w,p}$ is updated according to the service reliability of bus lines. The upper layer regenerates a new scheduling scheme and transmits it to the lower layer.

3.2. Formulation of the lower layer. This subsection describes the information of each part in the lower layer. The operation module is used to infer the state of the transit network. Then, some parameters such as the arrival time, the departure time, headway and the number of passengers per bus are forwarded into the training module. In the training module, the LSTM method is used to estimate the passenger demand in the future. Then, the information based on the LSTM is transmitted to the reinforcement learning module. Through the training module, the actions performed by different buses can be obtained. The symbols in the lower layer are shown in Table 3.

Table 3. Notation in the lower layer.

Parameters	Definition
$\lambda_i^{r,k}$	average passenger arrival rate at stop i in route r in period k
$q_i^{r,k}$	alighting rate of passengers at stop i in route r in period k
$t_{i,j}^{r,k}$	travel time between stops $j-1$ and j by bus i in route r in period k
Variables	**Definition**
$ta_{i,j}^{r,k}$	time instant at which bus i arrives at stop j in route r in period k
$ts_{i,j}^{r,k}$	service time of bus i at stop j in route r in period k
$td_{i,j}^{r,k}$	time instant at which bus i departs from stop j in route r in period k
$tz_{i,j}^{r,k}$	holding time of bus i at stop j in route r in period k
$A_{i,j}^{r,k}$	number of alighting customers of bus i at stop j in route r in period k
$B_{i,j}^{r,k}$	number of boarding customers of bus i at stop j in route r in period k
$\beta_{i,j}^{r,k}$	number of passengers unable to boarding the bus i at stop j due to the capacity constraints in route r in period k
$w_{i,j}^{r,k}$	number of passengers unable to boarding bus i at stop j due to acceleration strategy in route r in period k
$h_{i,j}^{r,k}$	headway between buses i and $i-1$ at stop j in route r in period k
$Q_{i,j}^{r,k}$	number of passengers in bus i when bus i departs from stop j in route r in period k
$tB_{m,i,j}^{r,k}$	time instant at which standby bus m is arranged after bus i from stop j to serve route r in period k
$Z_{i,j}^{r,k}$	binary, equals one if standby bus i provides service from stop j in route r in period k; otherwise, equals zero
$y_{i,j}^{r,k}$	binary, equals one if bus i arrives at stop j in route r in period k; otherwise, equals zero

3.2.1. Methodology of the operation module. The passenger demand is determined based on historical operation information. The arrival time of bus i at stop j is determined by the sum of the departure time at stop $j-1$ and the driving time between Stop $j-1$ and stop

j. We need to consider the departure time of bus $i-1$ at stop j to prohibit overtaking. Therefore, $ta^{r,k}_{i,j}$ can be determined by

$$ta^{r,k}_{i,j} = \max\{td^{r,k}_{i,j-1} + t^{r,k}_{i,j}, td^{r,k}_{i-1,j}\}. \qquad (7)$$

The departure time of bus i is related to the arrival time, service time, and the action performed by bus i at stop j. Here

$$td^{r,k}_{i,j} = ta^{r,k}_{i,j} + ts^{r,k}_{i,j} + tz^{r,k}_{i,j}. \qquad (8)$$

In this formulation, we can conclude that if the action performed by bus i at stop j is not holding, $tz^{r,k}_{i,j} = 0$. The service time of bus i at stop j is associated with the boarding and alighting process. In this paper, we assume that the boarding and alighting processes are carried out at the same time. Based on historical data, we find that the alighting process usually takes less time than the boarding processes. The times of boarding and alighting processes at stop j are linearly related to the number of passengers. The number of passengers getting off bus i at stop j can be expressed by

$$A^{r,k}_{i,j} = y^{r,k}_{i,j} \cdot q^{r,k}_j \cdot Q^{r,k}_{i,j-1} + (1 - y^{r,k}_{i,j+1}) \cdot q^{r,k}_{j+1} \cdot Q^{r,k}_{i,j-1}, \qquad (9)$$

$$y^{r,k}_{i,j} + y^{r,k}_{i,j+1} \geq 1. \qquad (10)$$

In the formulation (9), if $y^{r,k}_{i,j} = 0$ and $y^{r,k}_{i,j+1} = 1$, then bus i stops and provides service at stop $j+1$, and $A^{r,k}_{i,j} = 0$. Bus i skips stop j and passengers cannot get off the bus at stop j. If $y^{r,k}_{i,j} = 1$ and $y^{r,k}_{i,j+1} = 0$, bus i provides service at stop j and skips stop $j+1$. Passengers whose destination is stop $j+1$ need to get off bus i at stop j. If $y^{r,k}_{i,j} = 1$ and $y^{r,k}_{i,j+1} = 1$, then bus i does not perform the acceleration action and stops normally at two stops. Constraints (10) prevent the bus from skipping two consecutive stops. The bus can usually complete the scheduling task by skipping one stop at a time, and there is no need to skip multiple stops continuously to improve the service reliability of bus lines.

Besides, if the bus skipped multiple consecutive stops, it would cause great inconvenience to passengers and affect their travel experience. Passengers taking bus i at stop j in period k can be divided into three categories: the number of passengers $\beta^{r,k}_{i-1,j}$; the number of passengers $p^{1,r,k}_{i,j}$ who arrive at stop j during the period from bus $i-1$ leaving stop j to bus i arriving at stop j; the number of passengers $p^{2,r,k}_{i,j}$ who arrive at stop j during the period the bus i providing service at stop j. The service time at stop j depends on the larger value of the boarding and alighting times. The time required for passengers to board or alight is affected by the used space in the bus. The used space (expressed by $Q^{r,k}_{i,j-1}/C$) is an important indicator to measure the time required for passengers to boarding and alighting. We have

$$p^{1,r,k}_{i,j} = \lambda^{r,k}_i \cdot (ta^{r,k}_{i,j} - td^{r,k}_{i-1,j}). \qquad (11)$$

The service time $ts'^{r,k}_{i,j}$ at this stage could be expressed by

$$ts'^{r,k}_{i,j} = y^{r,k}_{i,j} \cdot c \cdot \max\{\delta \cdot (p^{1,r,k}_{i,j} + \beta^{r,k}_{i-1,j}), \sigma \cdot A^{r,k}_{i,j}\}. \qquad (12)$$

Here δ and σ are parameters representing the average time for passengers to boarding and alighting, respectively.

Define

$$\begin{cases} c = 1 & \text{if } 0 \leq Q^{r,k}_{i,j-1}/C \leq \tau, \\ c > 1 & \text{if } \tau < Q^{r,k}_{i,j-1}/C \leq 1 \end{cases} \qquad (13)$$

as a parameter which is related to the used space in bus i; τ represents the threshold of the used space. The passengers $p^{2,r,k}_{i,j}$ who arrive at stop j during the period $ts'^{r,k}_{i,j}$ could be defined as

$$p^{2,r,k}_{i,j} = \lambda^{r,k}_j \cdot ts'^{r,k}_{i,j}. \qquad (14)$$

Besides, based on historical data, the number of people getting off at stop j could be calculated. Thus, the number of passengers getting on $B^{r,k}_{i,j}$ can be expressed by

$$B^{r,k}_{i,j} = \begin{cases} y^{r,k}_{i,j} \cdot (p^{1,r,k}_{i,j} + \beta^{r,k}_{i-1,j} + p^{2,r,k}_{i,j} - w^{r,k}_{i,j}) \\ \quad \text{if } p^{1,r,k}_{i,j} + \beta^{r,k}_{i-1,j} + p^{2,r,k}_{i,j} \\ \quad \leq C - (Q^{r,k}_{i,j-1} - A^{r,k}_{i,j}), \\ y^{r,k}_{i,j} \cdot (C - (Q^{r,k}_{i,j-1} - A^{r,k}_{i,j}) - w^{r,k}_{i,j}) \\ \quad \text{otherwise}. \end{cases} \qquad (15)$$

Here $w^{r,k}_{i,j}$ could present the number of passengers unable to take bus i at stop j due to the acceleration action.

Existing studies have proved that the method of limiting the number of boarding passengers can effectively improve the quality of bus services (Vuchic, 1973; Larrain *et al.*, 2010). In this paper, the boarding limit action means that passengers at a stop are only allowed to get on the bus during the period of alighting. If the alighting process is over, the bus stops service and leaves the stop. Therefore,

$$w^{r,k}_{i,j} = \begin{cases} \max\{p^{1,r,k}_{i,j} + \beta^{r,k}_{i-1,j} + p^{2,r,k}_{i,j} - \sigma \cdot A^{r,k}_{i,j}/\delta, 0\} \\ \quad \text{if } p^{1,r,k}_{i,j} + \beta^{r,k}_{i-1,j} + p^{2,r,k}_{i,j} \\ \quad \leq C - (Q^{r,k}_{i,j-1} - A^{r,k}_{i,j}), \\ \max\{C - (Q^{r,k}_{i,j-1} - A^{r,k}_{i,j}) - \sigma \cdot A^{r,k}_{i,j}/\delta, 0\} \\ \quad \text{otherwise}. \end{cases} \qquad (16)$$

If the bus i performs the action of boarding limit, the value of $w^{r,k}_{i,j}$ needs to be calculated according to (16). Otherwise, $w^{r,k}_{i,j} = 0$. The service time of bus i at stop j could be obtained through

$$ts^{r,k}_{i,j} = y^{r,k}_{i,j} \cdot c \cdot \max\{\delta \cdot B^{r,k}_{i,j},\ \sigma \cdot A^{r,k}_{i,j}\}. \qquad (17)$$

The number of passengers on board at consecutive stops is

$$Q^{r,k}_{i,j} = Q^{r,k}_{i,j-1} + B^{r,k}_{i,j} - A^{r,k}_{i,j}. \qquad (18)$$

The number of passengers who cannot board due to capacity constraints or acceleration action is

$$\beta_{i,j}^{r,k} = \max\{p_{i,j}^{1,r,k} + \beta_{i-1,j}^{r,k} + \lambda_j^{r,k} \cdot ts_{i,j}^{r,k} - B_{i,j}^{r,k}, 0\}. \quad (19)$$

The headway between bus i and bus $i-1$ at stop j can be obtained by

$$h_{i,j}^{r,k} = ta_{i,j}^{r,k} - td_{i-1,j}^{r,k}. \quad (20)$$

The following constraints represent the time limitations for the standby bus m if this bus provides service at stop j in route r in period k,

$$td_{i,j}^{r,k} \le tB_{m,i,j}^{r,k} \le ta_{i+1,j}^{r,k}, \quad (21)$$

$$td_{m,j}^{r,k} \ge tB_{m,i,j}^{r,k} + ts_{m,j}^{r,k}. \quad (22)$$

In this paper, the service reliability of all stops is derived in the operation module. Related parameters are then transmitted to the training module. The service reliability of bus lines would be affected by the reliability of road traffic. Due to the complexity of road traffic, we use a macroscopic method (i.e., zone-based travel time reliability (ZTTR)) to evaluate the reliability of bus lines. Specifically, a well-known risk metric, i.e., the value at risk (VAR), is defined as follows:

$$\mathrm{VAR}(T) = \inf\{t : F(t) \ge \psi\}, \quad (23)$$

$$M = \frac{t - \mathrm{E}(t)}{\mathrm{E}(t)}, \quad (24)$$

where $\psi \in (0,1)$ is the confidence level, T denotes the random variable, t represent the travel time of passengers with a given OD pair, $F(t)$ indicates the cumulative distribution function of T, M means the normalized deviation of trips.

The conditional VAR (CVAR) to present the average of losses exceeding VAR is

$$\begin{aligned}\mathrm{CVAR}(M) &= \mathrm{E}[M | M \ge \mathrm{VAR}(M)] \\ &= \frac{1}{1-\psi} \int_{\mathrm{VAR}(M)}^{\infty} m f(m)\,\mathrm{d}m.\end{aligned} \quad (25)$$

The formulas

$$\begin{aligned} r_{i,j} &= \mathrm{VAR}(M) + \mathrm{E}[(M - \mathrm{VAR}(M))^+] \\ &= \int_{\mathrm{VAR}(M)}^{\infty} m f(m)\,\mathrm{d}m = (1-\psi)\mathrm{CVAR}(M), \end{aligned} \quad (26)$$

$$\mathrm{SR}_i = \sum_{j \in N} \mathrm{SR}_{i,j} = \sum_{j \in N} r_{i,j} \cdot wp_{i,j} \quad (27)$$

establish the relationship between the travel risk $r_{i,j}$ and ZTTR, where SR_i represents the ZTTR of zone i; $\mathrm{SR}_{i,j}$ indicates the ZTTR between zone i and zone j; $r_{i,j}$ denotes the average risk of passengers traveling from zone i to zone j; $wp_{i,j}$ is the weight of zone OD pair (i,j), usually expressed by the passenger demand. The service reliability between two stops in a bus line can be expressed by the ZTTR between two corresponding zones.

The improvement of service stability of bus stops is conducive to transferring some passenger demand to public transportation. Specifically, the demand transfer can be expressed as a utility function related to time and cost,

$$V_{ij}^{\mathrm{bus}} = a \cdot t_{ij}^{\mathrm{bus}} + b \cdot c_{ij}^{\mathrm{bus}}, \quad (28)$$

$$V_{ij}^{\mathrm{car}} = a \cdot t_{ij}^{\mathrm{car}} + b \cdot c_{ij}^{\mathrm{car}} + d, \quad (29)$$

$$P_{ij}^{\mathrm{bus}} = \frac{\exp(V_{ij}^{\mathrm{bus}})}{\exp(V_{ij}^{\mathrm{bus}}) + \exp(V_{ij}^{\mathrm{car}})}, \quad (30)$$

where V_{ij}^{bus} and V_{ij}^{car} are the utilities of travelling by bus and car, respectively; t_{ij}^{bus} and t_{ij}^{car} are the average travel time between zone OD pair (i,j) by bus and car, respectively; c_{ij}^{bus} and c_{ij}^{car} are the costs of travelling by bus and car, respectively; a, b and d are weight parameters; P_{ij}^{bus} represents the probability that passengers choose to travel by bus (i.e., the public transportation sharing rate). The value of parameter P_{ij}^{bus} can reflect the attractiveness of public transportation to passengers.

3.2.2. Methodology of the training module. The DRL method has excellent performance in dealing with the continuous decision-making problem. The DRL could achieve the optimal goal through continuous interaction with the environment. The time-varying characteristics of the scheduling problem in the transit network are significant. Therefore, this paper selects DRL to solve the proposed problem. Besides, the LSTM method is used to predict the service reliability of all bus stops. The future information would be integrated into the current state for decision making.

LSTM is a special kind of recurrent neural network (RNN) investigated to remember data over long durations. A conventional RNN also has a memory mechanism. However, the data stored in an RNN would increase exponentially after a large number of iterations. In contrast, LSTM has the ability to learn the long-term dependencies based on historical data (Cheng *et al.*, 2018). It is worth noticing that the gate structure in LSTM is crucial in processing historical data and storing information. Important information is stored and transferred to subsequent cells, while trivial data would be gradually forgotten.

In practice, the service reliability of bus stops is important for operators to formulate policies. Operators need to take measures to ensure the service reliability of stops seriously affected by disturbances. With the use of LSTM, historical operation data can be integrated into the current state. Specifically, each LSTM cell contains two elements: an observation unit and a state unit. An observation value o'_t could be obtained by inputting the state s_{t-1} and action a_{t-1} into the observation unit. The state unit integrates the historical data and the current observation value o_t. Then, the state unit maps these information to a new state s_t. State s_t would be transmitted to the next cell and get a new state s_{t+1}.

Based on historical data, the service reliability of bus stops in the next five time steps is predicted. The predicted data are mapped to the current state s_t. Then, the internal state s_t is forwarded to the DDQN cell for action selection.

The action value function of reinforcement learning is formulated as the expected value of the discounted reward, which is defined as

$$Q_t^{\pi}(s,a) = \mathrm{E}_{\pi}\Big[\sum_{k=0}^{\infty}\gamma^k \cdot r_{t+k}|s_t = s, a_t = a\Big]. \tag{31}$$

The quantity

$$Q^*(s,a) = \max\ Q^{\pi}(s,a) \tag{32}$$

denotes the optimal action value function. The goal of reinforcement learning is selecting an optimal policy to maximize the action value function. Here $\mathrm{E}(T)$ is the expectation of T; s_t denotes the state at time t; a_t represents the action selection at time t; γ is the discount rate; r represents the reward that can be obtained at time t; π denotes the policy that could be performed by the agent.

The DQN is a variant of reinforcement learning. The action selection and evaluation in the DQN algorithm are based on the same network, which may lead to overestimation. The objective function in the DQN can be expressed as

$$Y_t^{\mathrm{DQN}} = r_t + \gamma \cdot \max_a\ Q(s_{t+1}, a; \theta^-). \tag{33}$$

Van Hasselt *et al*. (2016) pointed out that using the same network parameter θ^- to select actions and evaluate Q values may lead to a nonzero lower limit of $Q^*(s,a)$. Specifically, in the DQN algorithm, the optimal Q value is updated by taking the average of the maximum values of all possible Q values. This optimal Q value thus would be greater than the maximum of the expected value of all Q values, which is the reason for the overestimation. To avoid this problem, Van Hasselt *et al*. (2016) proposed the DDQN algorithm. This method can effectively deal with complex problems based on historical data.

Two networks are used in the DDQN method. The first network is an online network, which is used to evaluate the current action value function. In this network, the parameter θ is used to represent the parameters in the neural network. The second network is the target network, which is used to calculate the value of the objective function Y_t. θ^- represents the parameters of the second neural network. The objective in the DDQN could be defined as

$$Y_t = r_t + \gamma \cdot Q(s_{t+1}, \arg\max_a\ Q(s_{t+1}, a; \theta); \theta^-). \tag{34}$$

The loss function represents the mean square error between the evaluated action value and the target value. The loss function is defined as

$$L(\theta) = \mathrm{E}(Y_t - Q(s_t, a_t, \theta))^2. \tag{35}$$

The evaluation network uses the method of gradient descent to minimize the loss function. The update method of θ is based on the gradient descent

$$\theta_{t+1} = \theta_t + \alpha_t \cdot \nabla_{\theta_t} \cdot L(\theta_t), \tag{36}$$

where α_t is the learning rate. In this algorithm, the parameter θ is used to select the action. The parameter θ^- is used to evaluate the selected action.

For the scheduling problem in the transit network, it is necessary to accurately define the state space, action space and reward function. These attributes are discussed below.

Definition of the state space. The state is used to describe the status of the transit network. The state of the transit network at time t can be expressed by $s_t = (L_i, Q_{i,j}^{r,k}, W_{i,j}^{r,t,k}, h_{i,j}^r)$, where

- L_i represents the location of bus i in the transit network;
- $Q_{i,j}^{r,k}$ indicates the number of passengers on board i when bus i departs from stop j in route r in period k;
- $W_{i,j}^{r,t,k}$ represents the number of waiting passengers when bus i arrives at stop j at time t; the parameter $W_{i,j}^{r,t,k}$ could be calculated via
$$W_{i,j}^{r,t,k} = p_{i,j}^{1,r,k} + \beta_{i-1,j}^{r,k}; \tag{37}$$
- $h_{i,j}^r$ represents the headway between buses i and bus $i-1$ at stop j in route r.

Definition of the action space. Action means the scheduling method used in the transit network. In this paper, five actions are mainly considered in the transit network: (i) skipping control stop; (ii) skipping next stop; (iii) boarding limit; (iv) holding; (v) adding standby buses into the bus fleet. The action can be selected according to the practical operation conditions and performed by bus i at time t.

Definition of the reward function. The conventional actions in the action set (i.e., skipping the control stop, skipping the next stop, boarding limit and holding) would have a variety of effects on passengers. For example, the skipping action could reduce the travel time of some passengers on board. But at the same time, this action may increase the waiting time of passengers at bus stops and the additional travel time of passengers whose destination is the skipped stop. The fifth action (i.e., adding standby buses into the bus fleet) would have a positive impact on the transit network, but the number of standby buses is limited. Therefore, it is necessary to make a reasonable plan to arrange standby buses.

The objective in this paper contains multiple parts, which is shown as follows. The waiting time of passengers at bus stops is

$$W_1 = \sum_{i\in V}\sum_{j\in N}\sum_{r\in R}\sum_{k\in K}(\xi\cdot\lambda_i^{r,k}\cdot(h_{i,j}^{r,k})^2 + \beta_{i-1,j}^{r,k}\cdot h_{i,j}^{r,k}), \quad (38)$$

where ξ is a coefficient. The in-vehicle time of passengers is

$$W_2 = \sum_{i\in V}\sum_{j\in N}\sum_{r\in R}\sum_{k\in K}(t_{i,j}^{r,k} + ts_{i,j}^{r,k})\cdot Q_{i,j-1}^{r,k} + tz_{i,j}^{r,k}\cdot Q_{i,j}^{r,k}. \quad (39)$$

The extra travel time of passengers due to the skipping strategy is

$$\begin{aligned}W_3 = \sum_{i\in V}\sum_{i'\in V}\sum_{j\in N}\sum_{r\in R}\sum_{k\in K}&(1-y_{i,j}^{r,k})\cdot q_j^{r,k}\cdot Q_{i,j-1}^{r,k}\\ &\cdot V(t_{i,j+1}^{r,k} + t_{i',N_r-j+1}^{r,k} + h_{i',N_r-j}^{r,k})\\ &+(1-y_{i,j+1}^{r,k})\cdot q_{j+1}^{r,k}\cdot Q_{i,j-1}^{r,k}\cdot h_{i+1,j}^{r,k}.\end{aligned} \quad (40)$$

The goal of this paper is to minimize the total travel time of passengers in the transit network, so three types of time (i.e., W_1, W_2, W_3) need to be considered comprehensively. The objective is defined as the total time with the state s,

$$T(s) = \alpha_1\cdot W_1 + \alpha_2\cdot W_2 + \alpha_3\cdot W_3. \quad (41)$$

where α_1, α_2, and α_3 are the coefficient values of different parts.

Here r is the reward that the agent could obtain after executing action a with state s. This indicator is used to measure the quality of the selected action. In this paper, we set

$$r = \begin{cases}\mu\cdot\dfrac{G-T(s')}{G} & \text{if } T(s') \text{ is lower with state } s' \text{ than } s,\\[2ex] \mu\cdot\dfrac{G-T(s')}{G}\cdot\eta & \text{otherwise.}\end{cases} \quad (42)$$

Here G has the same meaning in (4); it G is (infinite) positive number; μ ($\mu \geq 1$) could appropriately expand the reward value; η is a coefficient ($0 \leq \eta \leq 1$). This reward setting can encourage agents to find actions that can minimize the total travel time of passengers in the transit network. Besides, it could also reduce the exploration of non-optimal areas.

Algorithm 1. Training combining DDQN and LSTM.

Step 1. Initialize replay memory D.
Step 2. Initialize the parameter θ of the online network.
Step 3. Initialize the parameter $\theta' = \theta$ of the target network.
Step 4. for episode = $1 : E$ do
Step 5. Initialize the current state o_1.
Step 6. for time $t = 1 : T$ do
Step 7. get the initial state s_t based on the LSTM.
Step 8. with probability ε select a random action a_t.
Step 9. otherwise select an action $a_t = \arg\max_a Q(s_t; a_t; \theta)$
Step 10. Execute the action a_t, calculate the reward r_t, and collect the next observation o_{t+1}.
Step 11. Store the transition (s_t, a_t, r_t, s_{t+1}) into D.
Step 12. Sample a random mini-batch of transitions (s_k, a_k, r_k, s_{k+1}) from D.
Step 13. Set

$$Y_t = \begin{cases} r_j & \text{if episode terminates at step } j+1,\\ r_j + \gamma\cdot Q(s_{j+1}, \arg\max_a Q(s_{j+1}, a; \theta); \theta^-), \\ \quad \text{otherwise.}\end{cases}$$

Step 14. Calculate loss function based on Eqn. (35), and perform a gradient descent with Eqn. (36).
Step 15. $\theta^- = \zeta\cdot\theta + (1-\zeta)\cdot\theta^-$
Step 16. end for
Step 17. end for

3.3. Framework of the training module.

This paper proposes a novel DDQN algorithm to solve the scheduling problem in the transit network. The overall framework of the tailored algorithm is shown as Algorithm 1. In this algorithm, E represents the total number of episodes. Time t is defined as the time when buses arrive at stops or a standby bus needs to provide service. It is particularly noteworthy that the traditional DDQN method updates the parameter θ^- every C steps. However, it is difficult to determine an optimal value of C in practice. Therefore, in this algorithm, we update the value of parameter θ^- every iteration, as shown in Step 15 of the algorithm. The parameter ζ belongs to the internal $(0, 1)$.

4. Computational experiments

4.1. Background and data.

Experiments have been carried out in Beijing. The topological layout of all bus

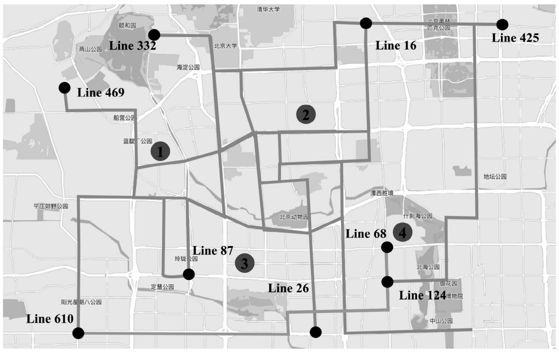


Fig. 2. Map of the Beijing transit network.

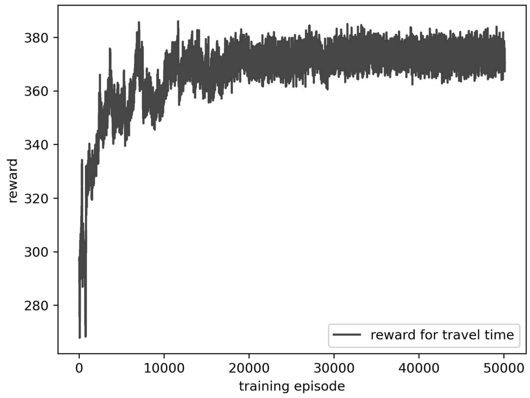


Fig. 3. Evolution of rewards in the training process.

lines is shown in Fig. 2. Table 4 gives detailed information of nine bus lines, including origin stops, terminal stops, length and the average headway during rush hours. The study period was from March 9, 2020 to March 13, 2020. We investigated the quality of bus services during the morning (i.e., 7 am–9 am) and evening rush hours (i.e., 5 pm–7 pm) every day.

Passengers are assumed to board at speed $\delta = 3$ s/person and alight at speed $\sigma = 1.5$ s/person. Here τ represents the threshold of the used space and its value in this paper is 0.8. In Eqn. (13), we assume that $c = 2$. The time span of each trip is 10 minutes. Buses in the investigated periods cruise at a speed of 20 km/h. Besides, we assume that the number of standby buses is 10. There are four standby locations in the research area, which are marked as big dots in Fig. 2. Standby buses are allowed to provide services for the bus line with poor service reliability. These buses would return to the original depot after a single service.

It is worth noting that the number of passengers arriving at bus stops during rush hours are obeyed by the Poisson distribution. Besides, the impact of some special events on bus services is not considered in this paper, such as the struggle between passengers.

4.2. Analysis of the results.

4.2.1. Reduction in passenger travel time. The proposed DDQN and LSTM was coded in Python. The algorithm was tested on a computer with an Intel ® Core (TM) i7–10700 CPU @ 2.90GHZ, 16GB RAM. As depicted in Fig. 3, the proposed algorithm is trained for 50000 episodes to learn an optimal policy. It was found that the rewards fluctuated rapidly in the first 10000 episodes, and gradually reached steady state in the next 10000 episodes. To verify the superiority of the proposed algorithm, it is set that the passenger demand fluctuates severely due to disturbances in the 20000 episodes of the experiment. From Figure 3, we can find that the reward value tends to stabilize after slight fluctuations. The stable state indicates that an optimal policy is found.

The total travel time of passengers for different scheduling strategies is calculated. The scheduling strategies are divided into no-control (NC), conventional control (CC) and hybrid control (HC) strategies. CC includes four actions: skipping the current stop, skipping the next stop, a boarding limit and holding. The HC adds the action of adding standby buses into bus fleets based on CC. The proposed TCM method in Section 3 could judge the bus stop whose service reliability might be very poor in the future. Standby buses can be used to provide temporary services, which is conducive to improving the service reliability of the specified line. In particular, if standby buses remain idle in the standby locations for a long time, it offers a good opportunity for re-using these buses as regular buses once in a while. Therefore, the travel experience of passengers may be improved theoretically, while operators can also make better use of idle resources.

The total travel time of passengers includes waiting time at bus stops, driving time and extra travel time caused by various actions. In fact, the speed of buses is constant and all scheduling strategies in this paper cannot reduce the driving time. However, the waiting time and exact travel time of passengers could be optimized. The total travel time of passengers per bus line under different scheduling strategies is shown in Fig. 4. We can observe that both the CC and HC have a positive effect on improving the quality of public transportation services. Compared with the inherent network, the total travel time of passengers for CC is roughly reduced by 40–50%. This proportion in the transit network based on HC roughly exceeds 60%. Furthermore, in order to verify the excellent performance of HC in energy conservation, the public transportation sharing rate and the energy consumption caused by private cars and buses are calculated (Hao

Table 4. Characteristics of the nine bus lines.

Line	Origin stop	Terminal stop	Length (km)	Headway during rush hours
16	ErLiZhuang	Zoo stop	12	11
26	ErLiZhuang	TianNingSiDong	23	5
68	MaGuanYingXi	ChangQiao road	12.7	10
87	DingHuiSiDong	JinMenQiao	15.5	15
124	DaTunDong	XiSi street	12.4	8
332	BeiGongMen stop	QianMen	21.1	8
425	NanWu	XiaoYing	19.4	8
469	The Summer Palace	WuLuJu	11	10
610	LuGu	XiYuan	20	9

et al., 2019; Goeke and Schneider, 2015). Parameters used in this paper are consistent with these two papers. Results show that compared with the transit network based on NC, the energy burned per passenger can be reduced by 15.67% if HC is performed to solve the dynamic scheduling problems. Obviously, HC is more suitable for solving the dynamic scheduling problem in large transit networks. This phenomenon explicitly shows that for a large transit network with numerous bus lines, the proposed strategy is able to outperform the existing control methods via providing services by standby buses.

The origin stop of Lines 16 and 26 is ErLiZhuang. Numerous passengers choose to travel by buses on these two lines. The total travel time of passengers on these two lines is thus high. A similar phenomenon also occurs in Lines 469 and 610, where buses begin to provide service from the depot or the famous tourist area. Interestingly, it is found that the total travel time of passengers on Line 87 is smallest among all bus lines. One of the reasons for this situation is that the arrival rate of passengers at bus stops on Line 87 is small. The other reason is that the OD of passengers is mainly concentrated in the middle area. The above two reasons lead to a low pressure of bus services on Line 87 during rush hours.

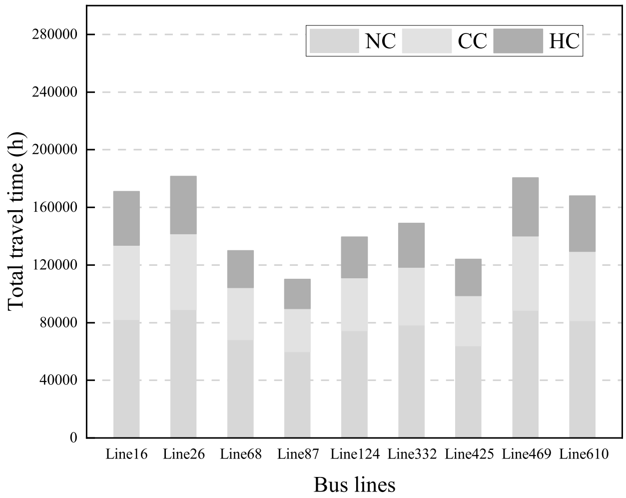


Fig. 4. Passenger travel time for different scheduling strategies.

4.2.2. Frequency of different actions. The frequency of various actions performed by buses during the research period are shown in Table 5. It can be found that the frequency of holding action exceeds that of the acceleration actions (i.e., skipping the current stop, skipping the next stop and the boarding limit). The reason for this phenomenon is that the holding action is simple with a low negative impact. The waiting time caused by holding is usually short, which is more acceptable for passengers than being skipped at a destination or the inability to take the current bus. Therefore, passengers are willing to accept a holding action different from acceleration actions. If it is necessary to use the skipping action to improve the service reliability of bus lines, the controlled bus tends to skip the next stop. This is because this action could reduce the exact travel time of passengers compared with skipping the current stop. If the next stop is skipped, the passengers with the destination of the next stop can get off at the current control stop. The additional travel time required is only the waiting time. In contrast, if the current stop is skipped, the passengers with the destination of the current stop have to get off at the next stop, the extra travel time of passengers includes the waiting time and twice the running time between the current stop and its next stop. For example, if bus i skips the control stop j, passengers need to alight at stop $j + 1$. Then these passengers will wait at stop $N - j$ and board bus i' back to stop $N - j + 1$. On the other hand, if bus i skips stop $j+1$, passengers whose destination is stop $j+1$ need to alight at stop j. Then they will wait for bus $i + 1$ to go to stop $j + 1$. This is because if the next stop is skipped, the additional travel time required for passengers is the waiting time at the current stop.

The scheduling results are shown in Fig. 5. In this figure, some bus trajectories in one day are displayed. The x-axis presents the rush hours, the y-axis indicates the ID of buses from #1 to #18 and the right y-axis shows the nine bus lines. Specifically, two solid lines are

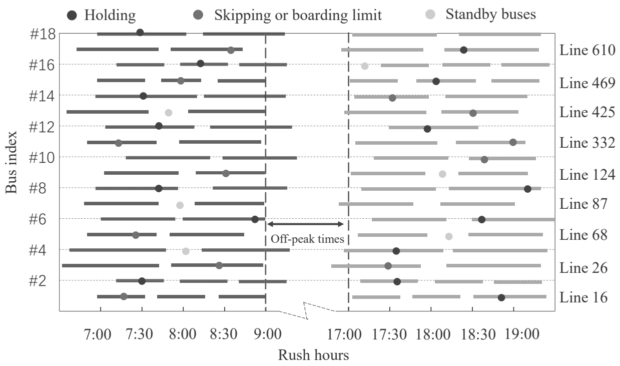


Fig. 5. Scheduling results.

used to indicate the operating conditions of these buses during the morning and evening rush hours, respectively. Several actions are marked by grey dots of varying scale. The period between 9:00 and 17:00 regards off-peak times separated by two dashed lines. The bus trajectory during this period is not of primary interest in this paper. As depicted in Fig. 5, the holding action is used more frequently than the acceleration action (i.e., skipping and boarding limit) when the headway of the bus fleet is unreasonable. Furthermore, it is found that these actions are concentrated between 7 am and 8 am during morning rush hours, while strategies are mainly applied to adjust the headway from 18:00 to 19:00 during evening rush hours.

In Fig. 6, the number of standby buses at different locations is shown. The heatmap on the left side of Fig. 6 represents the allocation of standby buses to different locations. The horizontal axis indicates the bus ID and the vertical axis indicates the date. The number in each box represents the location to which the bus is assigned. For example, the first square in the bottom left corner indicates that standby bus 1 is assigned to location 1 on Friday. The heatmap on the right side of Fig. 6 shows the frequency of bus services. The box indicates the number of trips performed by standby buses on that day. For example, the first square in the bottom left corner indicates that standby bus 1 completes a total of three trips on Friday. The service reliability of bus lines covered by locations 1 and 2 is poor. The number of services provided by buses at these locations is high. Through a large amount of simulations, we find that standby buses in location 1 usually serve lines 332 and 469. The bus placed in the standby location 2 serves passengers on bus lines 16 and 425. Buses in location 3 tend to serve passengers on lines 26, 87 and 610; buses in location 4 mainly provide services for bus lines 68 and 124. The comparison results of different strategies is going to be shown in the following subsection.

4.2.3. Distribution of headway. In this part, the distribution of headway during the investigated period for different strategies is shown in Fig. 7. The headway of all buses is determined every fifteen minutes. From Fig. 7, we can see that the headway of the original transit network is large and the service reliability is poor. The average headway of buses during the study period is about 9.21 minutes. Passengers need to spend a long time in the inherent transit network. The black line shows the results of the transit network based on the CC. The average headway is about 8.15 minutes. However, the frequency of large headway (e.g., greater than 10 minutes) are still very high. In contrast, HC has the best headway reduction performance. The average headway can be reduced to 6.47 minutes. The distribution of headway is mainly concentrated in the period between 5 and 8 minutes minutes. The frequency of large headway is greatly reduced compared with the NC and the CC. The reason for this phenomenon is that the CC only takes measurement for control after the disturbance causes serious congestion, which results in a delayed decision. In contrast, the proactive control method (i.e., the HC) can reduce the average headway due to the mechanism of controlling before disturbances.

The headway distribution for different control strategies is summarized in Table 6. The headway in this paper is divided into three levels: 0–6 minutes, 6–10 minutes, and more than 10 minutes. From Table 6, we can find that CC has a weak effect on improving the quality of bus services. Compared with the inherent network, the HC can significantly increase the proportion of headway in 0–6 minutes. Applying the HC to solve the dynamic scheduling problem is conducive to reducing the proportion of headway exceeding ten minutes. This is because the standby bus strategy would give priority to assigning buses to serve passengers on bus lines with poor service reliability. The optimization of the headway of 6–10 minutes is not obvious, which is limited by the number of standby buses.

5. Conclusions

In this paper, standby buses are considered to provide temporary or regular services for some bus lines. Besides, it is difficult to determine when to take action due to complex traffic conditions. We proposed a two-layer control method to solve the scheduling problems in the transit network. The optimization objective is to minimize the total travel time of passengers in the transit network. The simulation results show that our method can effectively improve the service reliability of the transit network. Our findings are the following:

1. The two-layer control method is an effective real-time scheduling method. It is very important

Table 5. Frequency of different actions for each bus line during the investigated period.

Action	Line/frequency								
	Line 16	Line 26	Line 68	Line 87	Line 124	Line 332	Line 425	Line 469	Line 610
Skipping the current stop	2	1	2	4	3	6	1	2	4
Skipping the next stop	10	15	12	11	19	8	10	14	16
Boarding limit	22	10	19	18	13	11	12	17	8
Holding	62	58	70	68	76	82	64	78	85
Standby bus	16	13	4	7	5	19	12	29	13
No-control	110	132	145	102	137	141	144	165	134

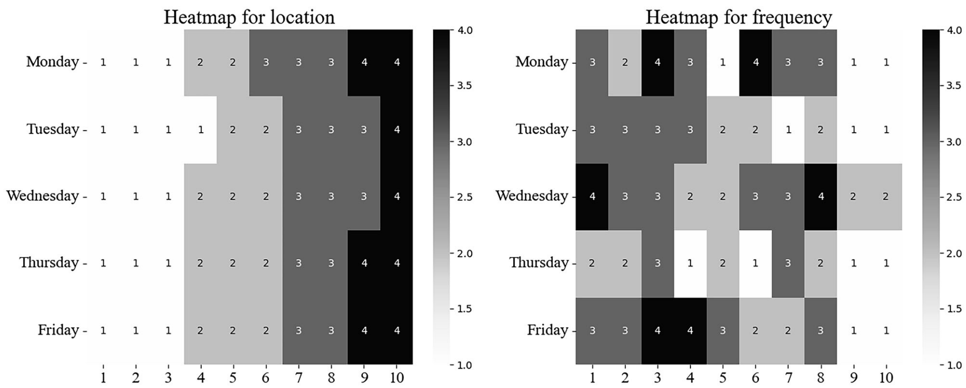


Fig. 6. Location and service frequency of standby buses during the research period.

to accurately describe the environmental state, such as headway, service reliability and driving time. These parameters have significant impact on the optimization results of the tailored DDQN algorithm.

2. Passengers are willing to accept a holding action other than skipping actions. The average frequency of the action of skipping the next stop is higher than that of skipping the current stop. This is because skipping the next stop would lead to a less exact travel time for passengers.

3. HC could further improve the service reliability of the transit network on the basis of CC. The standby bus strategy would give priority to assigning buses to serve passengers on bus lines with poor service reliability. Our results show that HC benefits from placing standby buses at the designated location in advance, and outperforms CC in terms of reducing passenger travel time and the average headway of bus fleets. It is conducive to the long-term development of the bus services.

Acknowledgment

This research was supported by National Key R&D Program of China (2018YFB1600503), the National Natural Science Foundation of China (U1811463 and 51878019), the Fundamental Research Funds for the Central Universities, and the China Postdoctoral Science Foundation (2022M710308).

References

Alesiani, F., Moreira-Matias, L. and Faizrahnemoon, M. (2018). On learning from inaccurate and incomplete traffic flow data, *IEEE Transactions on Intelligent Transportation Systems* **19**(11): 3698–3708.

Argote-Cabanero, J., Daganzo, C.F. and Lynn, J.W. (2015). Dynamic control of complex transit systems, *Transportation Research B: Methodological* **81**: 146–160.

Aslani, M., Mesgari, M.S. and Wiering, M. (2017). Adaptive traffic signal control with actor-critic methods in a real-world traffic network with different traffic disruption events, *Transportation Research C: Emerging Technologies* **85**: 732–752.

Barnett, A. (1974). On controlling randomness in transit operations, *Transportation Science* **8**(2): 102–116.

Cao, S., Shao, H. and Shao, F. (2022). Sensor location for travel time estimation based on the user equilibrium principle: Application of linear equations, *International Journal of Applied Mathematics and Computer Science* **32**(1): 23–33, DOI: 10.34768/amcs-2022-0003.

Ceder, A. (1988). Designing transit short-turn trips with the elimination of imbalanced loads, *in* J.R. Daduna *et al.* (Eds), *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems, Vol. 308, Springer, Berlin, pp. 288–303.

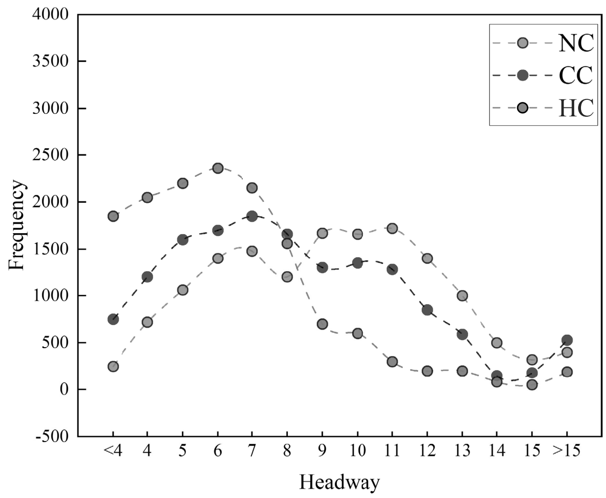


Fig. 7. Distribution of headway for different strategies.

Table 6. Headway distribution for different control strategies.

Strategy	Headway distribution (%)		
	0–6 minutes	6–10 minutes	>10 minutes
NC	23.2	40.7	36.1
CC	35.0	41.1	23.9
HC	58.4	34.5	7.1

Ceder, A. and Stern, H.I. (1981). Deficit function bus scheduling with deadheading trip insertions for fleet size reduction, *Transportation Science* **15**(4): 338–363.

Cheng, Y., Huang, Y., Pang, B. and Zhang, W. (2018). ThermalNet: A deep reinforcement learning-based combustion optimization system for coal-fired boiler, *Engineering Applications of Artificial Intelligence* **74**: 303–311.

Cortés, C.E., Jara-Díaz, S. and Tirachini, A. (2011). Integrating short turning and deadheading in the optimization of transit services, *Transportation Research A: Policy and Practice* **45**(5): 419–434.

Daganzo, C.F. and Pilachowski, J. (2011). Reducing bunching with bus-to-bus cooperation, *Transportation Research B: Methodological* **45**(1): 267–277.

Estrada, M., Mensión, J., Aymamí, J.M. and Torres, L. (2016). Bus control strategies in corridors with signalized intersections, *Transportation Research C: Emerging Technologies* **71**: 500–520.

Fang, K., Fan, J. and Yu, B. (2022). A trip-based network travel risk: Definition and prediction, *Annals of Operations Research*, DOI: 10.1007/s10479-022-04630-6.

Farazi, N.P., Zou, B., Ahamed, T. and Barua, L. (2021). Deep reinforcement learning in transportation research: A review, *Transportation Research Interdisciplinary Perspectives* **11**: 100425.

Fu, L. and Yang, X. (2002). Design and implementation of bus-holding control strategies with real-time information, *Transportation Research Record* **1791**(1): 6–12.

Furth, P.G. (1985). Alternating deadheading in bus route operations, *Transportation Science* **19**(1): 13–28.

Furth, P.G. (1986). Zonal route design for transit corridors, *Transportation Science* **20**(1): 1–12.

Goeke, D. and Schneider, M. (2015). Routing a mixed fleet of electric and conventional vehicles, *European Journal of Operational Research* **245**(1): 81–99.

Guan, Y., Li, S.E., Duan, J., Li, J., Ren, Y., Sun, Q. and Cheng, B. (2021). Direct and indirect reinforcement learning, *International Journal of Intelligent Systems* **36**(8): 4439–4467.

Hall, R., Dessouky, M. and Lu, Q. (2001). Optimal holding times at transfer stations, *Computers & Industrial Engineering* **40**(4): 379–397.

Hao, J., Liu, X., Shen, X. and Feng, N. (2019). Bilevel programming model of urban public transport network under fairness constraints, *Discrete Dynamics in Nature and Society* **2019**: 2930502.

Hu, H., Jia, X., He, Q., Fu, S. and Liu, K. (2020). Deep reinforcement learning based AGVs real-time scheduling with mixed rule for flexible shop floor in Industry 4.0, *Computers & Industrial Engineering* **149**: 106749.

Jordan, W.C. and Turnquist, M.A. (1979). Zone scheduling of bus routes to improve service reliability, *Transportation science* **13**(3): 242–268.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks, *Communications of the ACM* **60**(6): 84–90.

Larrain, H., Giesen, R. and Muñoz, J.C. (2010). Choosing the right express services for bus corridor with capacity restrictions, *Transportation Research Record* **2197**(1): 63–70.

Laskaris, G., Cats, O., Jenelius, E., Rinaldi, M. and Viti, F. (2019). Multiline holding based control for lines merging to a shared transit corridor, *Transportmetrica B: Transport Dynamics* **7**(1): 1062–1095.

Li, L., Lv, Y. and Wang, F.-Y. (2016). Traffic signal timing via deep reinforcement learning, *IEEE/CAA Journal of Automatica Sinica* **3**(3): 247–254.

Luo, S. (2020). Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning, *Applied Soft Computing* **91**: 106208.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. and Hassabis, D. (2015). Human-level control through deep reinforcement learning, *Nature* **518**(7540): 529–533.

Petit, A., Lei, C. and Ouyang, Y. (2019). Multiline bus bunching control via vehicle substitution, *Transportation Research B: Methodological* **126**: 68–86.

Petit, A., Ouyang, Y. and Lei, C. (2018). Dynamic bus substitution strategy for bunching intervention, *Transportation Research B: Methodological* **115**: 1–16.

Van Hasselt, H., Guez, A. and Silver, D. (2016). Deep reinforcement learning with double q-learning, *Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, USA*, pp. 2094–2100.

Van Oort, N., Boterman, J. and van Nes, R. (2012). The impact of scheduling on service reliability: Trip-time determination and holding points in long-headway services, *Public Transport* **4**(1): 39–56.

Vuchic, V.R. (1973). Skip-stop operation as a method for transit speed increase, *Traffic Quarterly* **27**(2): 307–327.

Wang, J. and Sun, L. (2020). Dynamic holding control to avoid bus bunching: A multi-agent deep reinforcement learning framework, *Transportation Research C: Emerging Technologies* **116**: 102661.

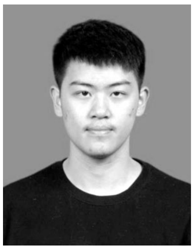
Wang, P. and Chan, C.-Y. (2017). Formulation of deep reinforcement learning architecture toward autonomous driving for on-ramp merge, *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan,* pp. 1–6.

Wang, T. and Tang, Y. (2021). Comprehensive evaluation of power flow and adjustment method to restore solvability based on GCRNN and DDQN, *International Journal of Electrical Power & Energy Systems* **133**: 107160.

Wang, W., Chen, X., Musial, J. and Blazewicz, J. (2020). Two meta-heuristic algorithms for scheduling on unrelated machines with the late work criterion, *International Journal of Applied Mathematics and Computer Science* **30**(3): 573–584, DOI: 10.34768/amcs-2020-0042.

Xuan, Y., Argote, J. and Daganzo, C.F. (2011). Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis, *Transportation Research B: Methodological* **45**(10): 1831–1845.

Zhu, Y., Mao, B., Bai, Y. and Chen, S. (2017). A bi-level model for single-line rail timetable design with consideration of demand and capacity, *Transportation Research C: Emerging Technologies* **85**: 211–233.



Zhongshan Liu is an MSc student at the School of Transportation Science and Engineering, Beihang University. His current research interests include intelligent transportation systems, traffic assignment and deep reinforcement learning technology.

Bin Yu received his PhD degree at the Dalian University of Technology in 2006. Now, he is a professor at the School of Transportation Science and Engineering, Beihang University. His research activity concentrates on the development and application of innovative models and ITS tools. He has published more than 60 papers in SCI/SSCI indexed journals.

Li Zhang is a PhD student at the School of Transportation Science and Engineering, Beihang University. His current research interests include intelligent transportation systems, traffic assignment and big data for transportation.

Wensi Wang is a postdoctoral researcher in the School of Transportation Science and Engineering, Beihang University. Her current research interests include intelligent transportation systems, public transit control and machine learning technology.

Received: 21 December 2021
Revised: 4 May 2022
Re-revised: 17 July 2022
Accepted: 27 July 2022