Integrated optimization of timetable, bus formation, and vehicle scheduling in autonomous modular public transport systems

Liu, Zhengke; Homem de Almeida Correia, Gonçalo; Ma, Zhenliang; Li, Shen; Ma, Xiaolei

DOI
10.1016/j.trc.2023.104306

Publication date
2023

Document Version
Final published version

Published in
Transportation Research Part C: Emerging Technologies

Transportation Research Part C 155 (2023) 104306

Contents lists available at ScienceDirect

Transportation Research Part C

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

Integrated optimization of timetable, bus formation, and vehicle scheduling in autonomous modular public transport systems

Zhengke Liu [a], Gonçalo Homem de Almeida Correia [b], Zhenliang Ma [c], Shen Li [d], Xiaolei Ma [a,e,*]

[a] School of Transportation Science and Engineering, Beihang University, Beijing 100191, China
[b] Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628 CN Delft, the Netherlands
[c] Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Stockholm 10044, Sweden
[d] Department of Civil Engineering, Tsinghua University, Beijing 10084, China
[e] Key Laboratory of Intelligent Transportation Technology and System, Ministry of Education, Beijing 100191, China

ARTICLE INFO

Keywords:
Public transport
Autonomous modular vehicle
Timetable
Flexible vehicle scheduling
Integrated optimization
Alternating Direction Method of Multipliers

ABSTRACT

This paper presents a joint optimization of the timetable, bus formation, and vehicle scheduling in a flexible public transport (PT) system that utilizes autonomous modular vehicles (AMVs). In this system, AMVs have the capability to detach or join with each other at intermediate stops along the route to dynamically adjust the bus formation (capacity). To increase vehicle utilization, a flexible scheduling strategy is proposed that allows AMVs to detach from one modular bus and join another modular bus in either direction of a bidirectional line. In particular, the penalty cost for each detachment or joining operation, as well as the limited number of available AMVs is explicitly considered. We formulate a unified model for the integrated optimization of the modular bus service (timetable and bus formation) and vehicle scheduling by introducing two types of decision variables. The objective is to minimize overall system costs, including passenger waiting time cost, operational costs, and detachment/joining penalty costs. The two types of decision variables are coupled by a vehicle resource consistency constraint, ensuring the conformity of the modular bus service and vehicle scheduling decisions. To tackle the complexity of our model, the Alternating Direction Method of Multipliers (ADMM) is employed to decompose it into two subproblems, which can be efficiently solved using a customized forward dynamic programming algorithm and a commercial solver. The model is validated using illustrative examples and a real-world instance from the Beijing Public Transport system, and it is compared with two benchmark models. Our results demonstrate the efficiency of the ADMM-based solution framework for solving the integrated optimization model. Furthermore, our findings indicate that the use of AMVs in PT systems can lead to reduced overall system costs and increased vehicle utilization.

* Corresponding author at: School of Transportation Science and Engineering, Beihang University, Beijing 100191, China.
E-mail address: xiaolei@buaa.edu.cn (X. Ma).

https://doi.org/10.1016/j.trc.2023.104306
Received 1 October 2022; Received in revised form 9 August 2023; Accepted 13 August 2023
Available online 21 August 2023

1. Introduction

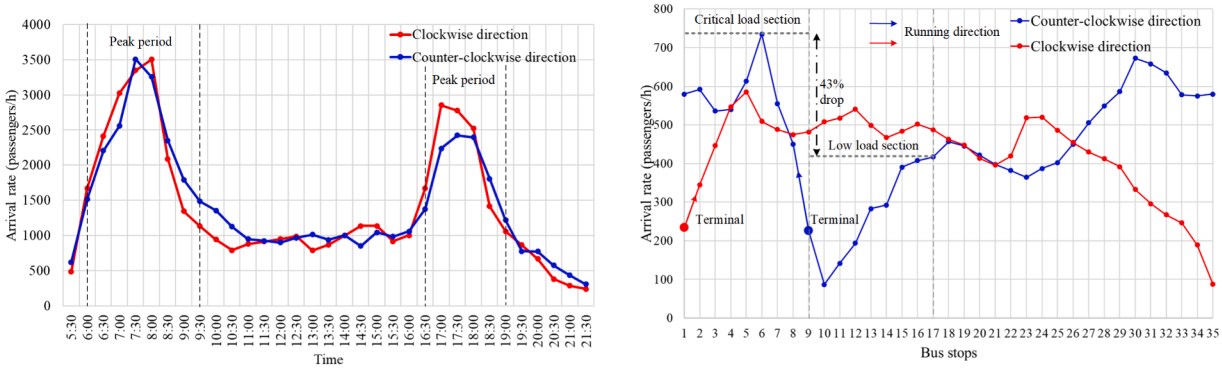
1.1. Motivation

Passenger demand shows strong temporal and spatial fluctuations over a typical weekday in many public transport (PT) systems. Fig. 1 illustrates the temporal and spatial patterns of passenger demand based on Automatic Fare Collection (AFC) transaction data collected from Line 300, a loop bus line in Beijing, on October 8, 2019 (a weekday). As demonstrated in Fig. 1(a), passenger demand surges during peak periods but remains at a lower level during off-peak periods. This phenomenon, caused by commuter trip patterns, often results in an imbalance between passenger demand and transportation supply, leading to excessive passenger waiting time and an increase in vehicle operational costs (Chen et al., 2019). To address these issues, peak/off-peak-based or demand-driven dispatching strategies have been applied by operators and have attracted the attention of many scholars (Niu and Zhou, 2013; Sun et al., 2014; Zhuo et al., 2023). However, these two dispatching strategies have identical drawbacks when dealing with significant temporal demand heterogeneity. For instance, because the bus capacity is fixed in both strategies, the bus load during off-peak periods may be relatively low, resulting in significant energy waste due to hauling empty units. The spatial demand heterogeneity of the PT system can also cause the same issue due to the spatial coverage requirement for PT network design. As shown in Fig. 1(b), the average bus load can be significantly greater at certain sections of a bus line than at other sections. Therefore, if the bus dispatching headways are determined using the well-known maximum loading point method (Ceder, 2016), which ensures that the bus load at the busiest point on the entire route does not exceed the capacity of the bus, then buses will be highly underutilized for the remaining parts of PT routes.

The emerging technology of autonomous modular vehicles (AMVs) holds great potential for solving the aforementioned challenges of PT systems. As illustrated in Fig. 2, AMV technology utilizes identical modular bus units with a specific capacity that can physically detach and join with each other, resulting in a flexible capacity design. Notably, AMV technology enables in-motion transfer, where modular bus units can detach and join while in motion on roads, allowing passengers to transfer between modular bus units during travel. These distinctive features allow AMV-based PT systems to effectively adjust vehicle formation (capacity) and adopt flexible vehicle scheduling to respond to the heterogeneous spatiotemporal passenger demand. In summary, AMV-based PT systems have the potential to offer higher-quality services by enabling better utilization of vehicle resources, ultimately leading to a reduction in both operational costs and passenger waiting time cost.

Recently, researchers have started investigating the potential benefits of AMV-based PT systems and developing various operational strategies to enhance their efficiency and adaptability (Chen and Li, 2021; Chen et al., 2020b; Shi and Li, 2021). Fig. 3 illustrates three different operational strategies for a fixed-route PT system. The current strategy (Fig. 3(a)) uses fixed-capacity buses that can only adjust dispatching headway to match time-dependent passenger demand, resulting in high empty seat rates during off-peak periods. In Fig. 3(b), Shi and Li (2021) introduced AMVs into a corridor system, allowing the modular bus to adjust its formation at the terminal depot to optimize both dispatching headway and bus capacity. However, high empty seat rates of modular buses may still occur on sections with low passenger demand. To address this issue, Chen and Li (2021) further explored the potential of AMVs and proposed a fully flexible capacity operational strategy, as shown in Fig. 3(c). This strategy involves modular bus units that can detach and join with each other while in motion on roads, enabling more frequent dispatches of small capacity vehicles to match time-dependent passenger demand. As a result, both operational costs and passenger waiting time cost are reduced.

It is noteworthy that the current literature on AMV-based PT systems primarily focuses on the joint optimization of timetable and bus formation plan for unidirectional lines. However, it should be highlighted that the advantages of AMVs are not limited to their ability to adapt to time-dependent passenger demand during the timetabling stage. They also offer flexibility in vehicle resource utilization during the vehicle scheduling stage, which involves the assignment of buses to cover a specific set of timetabled services (Ji et al., 2021; Liu et al., 2020). While a fully flexible capacity operational strategy for one direction of a bidirectional line could result in an efficient timetable and a bus formation plan, modular bus units that detach from the modular bus along the direction may accumulate at intermediate depots due to the spatial heterogeneity of passenger demand. This presents a significant challenge to vehicle



(a) Time-dependent passenger arrival rate for bus line 300 over one weekday

(b) Average bus load per bus stop for bus line 300 from 6:00 am to 9:00 am

Fig. 1. Temporal and spatial patterns of passenger demand for bus line 300 in Beijing on October 8, 2019.

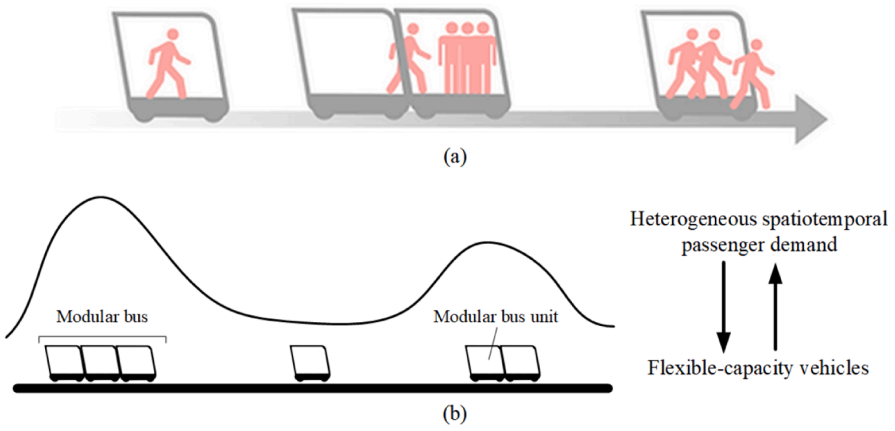


Fig. 2. Emerging AMV technology and usage in the PT system. (a) Autonomous modular vehicles (source: https://www.next-future-mobility.com/) and (b) Flexible capacity modular buses match the spatiotemporal passenger demand.

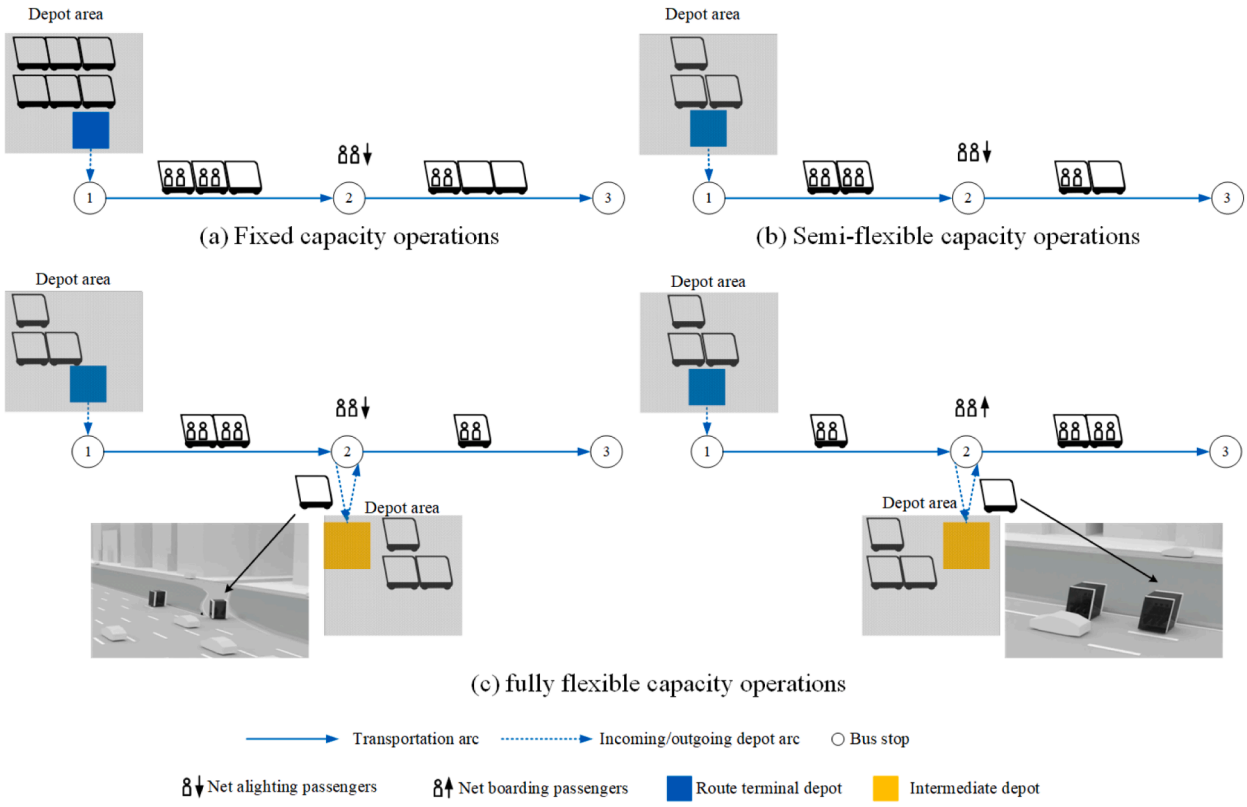


Fig. 3. Different operational strategies of PT system.

scheduling. One potential solution to this issue is to schedule vehicles in both directions on the line simultaneously and allow all available modular bus units at intermediate depots to join modular buses in either direction, ultimately resulting in increased utilization of modular bus units (vehicle resources). Besides, the detachment and joining operations in AMV-based PT system do not result in any additional time loss. However, there are inherent risks associated with physically detaching and joining modular bus units while maintaining precise speed control, and these risks should be taken into consideration. Additionally, whenever modular bus units are detached or joined, passengers may need to move around inside the modular bus units to accommodate the changes, which could

potentially cause discomfort or inconvenience for some passengers. Therefore, the frequency of detachment and joining operations should be considered when implementing AMVs into PT systems.

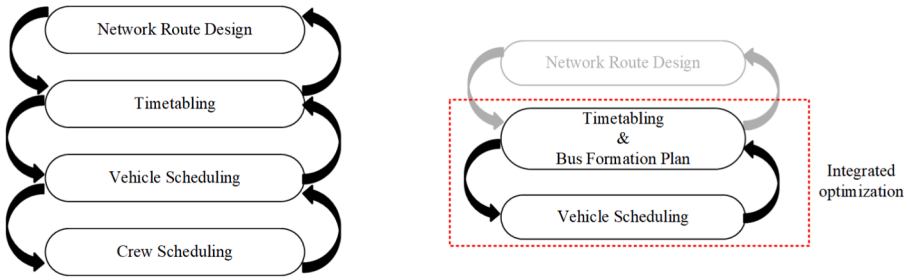
This study aims to formulate and solve the problem of jointly optimizing the timetable, bus formation, and vehicle scheduling for a bidirectional bus line considering en-route detachment and joining of autonomous modular bus units, with the objective of reducing both operational costs and passenger waiting time cost. Additionally, we introduced a penalty cost to account for both the risks associated with detachment and joining operations, as well as any potential discomfort that passengers may experience when moving from one modular bus unit to another. The strict vehicle capacity of the modular bus and the limited number of available modular bus units are explicitly considered. To further improve the flexibility of the operation, scheduling modular bus units in both directions simultaneously is incorporated into the integrated problem.

1.2. Related literature

1.2.1. Integration of timetable and vehicle scheduling

The design process of PT systems (Fig. 4(a)) is commonly divided into four stages to reduce computation complexity (Ceder, 2016). The first stage involves the network route design, which determines the alignment of the bus lines and the location of the bus stops. This strategic activity is infrequently carried out due to the potential for significant disruption to passengers resulting from substantial changes to the PT network. The final stage involves crew scheduling, which assigns drivers based on the results of vehicle scheduling. In case of driver shortages, vehicle scheduling must be coordinated with crew scheduling to ensure adequate staffing (Dai et al., 2023). However, the scope of this paper is limited to autonomous buses which do not involve significant human labor in daily operations (Stevens et al., 2022). Timetabling and vehicle scheduling are two intermediate stages in the overall planning process and are critical to ensuring high-quality service (Ceder and Wilson, 1986). Timetabling sets the departure times for all trips to minimize passenger waiting time, using either even headways (Shafahi and Khani, 2010; Yu et al., 2011) or uneven headways based on time-dependent passenger demand (Niu and Zhou, 2013; Shang et al., 2018; Sun et al., 2014). Meanwhile, vehicle scheduling assigns buses to trips to minimize the number of vehicles used and reduce operational costs. Typically, planning decisions in each stage are optimized sequentially, with solutions obtained in the previous stage potentially being suboptimal or even infeasible to implement in the subsequent stage (Carosi et al., 2019; Wu et al., 2022). Therefore, the integrated optimization of these two crucial stages has received significant attention from scholars. The literature can be divided into two categories based on research scope: network scale and single-line scale. Studies at the network scale aim to optimize the utilization of vehicle resources and improve transfer synchronization for passengers by integrating timetabling and vehicle scheduling across multiple lines in a transportation network (Fonseca et al., 2018; Liu and Ceder, 2018). In this study, we concentrate on the integration of timetabling and vehicle scheduling at the single-line scale, incorporating flexible scheduling tactics. Carosi et al. (2019) developed a multicommodity-flow model to optimize both timetable and vehicle scheduling for a complex bidirectional bus line with multiple patterns in one direction, corresponding to different terminal choices. Hu et al. (2023) introduced a method for integrating the generation of multitype bus timetables and the formation of chains of trips, considering the multiperiod characteristics of passenger flow on a bidirectional bus line. These studies have explored the advantages of integration problem, utilizing flexible scheduling tactics such as multiple trip patterns and vehicle types. However, it is important to note that oversaturated passenger demand is a common occurrence during peak periods in PT systems and should be explicitly considered (Niu and Zhou, 2013). Wu et al. (2022) proposed a mixed-integer linear programming model that integrates timetable and vehicle scheduling while also incorporating ride-matching for a loop shuttle bus line. The model considers multiple vehicle types and stop-skipping tactics. The integration of timetable and vehicle scheduling with short-turning strategy can provide more services in high-demand segments, thereby reducing the total number of detained passengers and the total waiting time of passengers at oversaturated stops (Yang et al., 2021; Yuan et al., 2022). However, it is worth noting that while stop-skipping and short-turning strategies can improve the efficiency of vehicle utilization, they may miss passengers at skipped stops and shortened routes, leading to inconvenience and reduced service reliability.

Our study presents an integrated model that optimizes timetable, vehicle scheduling, and bus formation simultaneously, while accounting for oversaturated conditions on a bidirectional line. Furthermore, a flexible vehicle scheduling strategy has been



(a) Current public transport hierarchical planning process

(b) AMV-based transport hierarchical planning process and research scope of this paper

Fig. 4. Different types of PT hierarchical planning processes and research scope of this paper.

introduced, allowing modular bus units to detach from a modular bus and join another modular bus in either direction. This strategy ensures that the original modular bus can continue its route without any delays.

1.2.2. Designing flexible capacity for transit services

With the advancement of AMV technology, modular bus units that can be physically detached and joined offer significant potential for designing transit services with flexible capacity to cater to passenger demand and reduce operational costs. Therefore, it is worth exploring the related literature on flexible capacity design for transit services. One significant approach is platooning whereby vehicles drive closely together as a platoon, facilitated by semi-autonomous driving technologies such as Cooperative Adaptive Cruise Control (CACC). In a platoon, the lead vehicle can be either manually driven or autonomous, while the following vehicles are equipped with autonomous driving systems. They are linked through wireless communication technology (Bhoopalam et al., 2018; Wang et al., 2022). A recent simulation-based study by Repoux et al. (2021) analyzed an on-demand passenger transportation service that employs platooning. The proposed service allows trailers following the lead vehicle to detach from the platoon and travel autonomously to nearby stations for picking up and dropping off passengers. The platoon can then continue its route without any delay. Moreover, platooning is potentially beneficial in freight transportation applications by providing flexible transportation capacities that match demand and save costs (Scherr et al., 2022; Scherr et al., 2019; Völz et al., 2022). One notable operational difference between platooned vehicles and AMVs is that platooned vehicles are virtually coupled with a small gap, making passenger movement between vehicles impossible. Chen et al. (2019) firstly introduced AMV technology into the PT system and developed an integrated optimization model for a shuttle system's timetable and bus formation problem. Dai et al. (2020) focused on the mixed traffic with AMVs and Human-driven Buses (HBs), and provided a joint optimization method for determining the appropriate dispatching time and bus type (HBs or AMVs). Chen and Li (2021) and Chen et al. (2022) employed both discrete modeling and continuum approximation methods to investigate the integrated design of timetable and flexible bus formation, while considering station-wise docking in a transit corridor. Based on them, Tian et al. (2023) further incorporated the modular bus units reposition process and the associated operation costs. Recently, Khan et al. (2023) analyzed the application of AMV technology to mitigate bus bunching using simulation. Note that our paper focuses on the application of AMV technology to fixed-route public transit services. The additional flexibility of AMV technology has applications in a wide range of transportation modes, including emergency vehicle services, on-demand transit, and ride-sharing. Interested readers could refer to the relevant literature in (Dakic et al., 2021; Gong et al., 2021; Hannoun and Menéndez, 2022; Liu et al., 2021a; Wu et al., 2021).

Prior research has considered the utilization of AMV technology to propose a range of operational strategies for improving the flexibility of PT systems. A significant limitation of these studies is their tendency to focus on a single stage of the hierarchical planning process. Only a few studies, including those by Ji et al. (2021), Chen et al. (2022), and Tian et al. (2023), have considered the integrated optimization of timetable, bus formation, and vehicle scheduling. However, Ji et al. (2021) did not account for oversaturation conditions or the en-route detachment and joining of modular bus units, which limits the practicality and flexibility of the system. Chen et al. (2022) and Tian et al. (2023) included the modular bus unit repositioning process in the integrated problem within a transit corridor, but only in one direction. In our study, we simultaneously schedule modular bus units in both directions of a bidirectional line to increase vehicle utilization. This involved allowing modular bus units to detach from a modular bus and join another modular bus in either direction. Additionally, it is important to acknowledge that the physical detachment and joining of AMVs may not be entirely risk-free, and the in-motion transfer could potentially cause discomfort or inconvenience for some passengers. The current studies lack consideration of the associated costs of physical detachment and joining operations.

1.3. Contributions

To fill the above gap, this paper proposes an integrated optimization of timetable, bus formation, and vehicle scheduling in a flexible AMV-based PT system. The key contributions are:

- Constructing a unified model based on a space–time-state (STS) network and a time–space network to integrate the timetable, bus formation and vehicle scheduling and minimize the total passenger waiting time cost, operational costs and penalty costs of detachment and joining operations.
- Introducing a novel vehicle scheduling strategy allowing modular bus units to detach from a modular bus and join another modular bus in either direction to increase the flexibility of vehicle scheduling.
- Proposing a solution framework based on Alternating Direction Method of Multipliers (ADMM) to decompose the integrated model into two subproblems: the direction-specific timetable and bus formation plan subproblem, and the vehicle scheduling subproblem. To solve the direction-specific subproblems efficiently, we use a customized dynamic programming (DP) algorithm, while the vehicle scheduling subproblem is solved using a commercial solver to obtain both primal and dual solutions.
- Validating the performance of our proposed model and algorithm using a real-world case study. Comparing our approach with two benchmark operational strategies and with state-of-the-art commercial solvers such as Gurobi.

The remainder of this paper is structured as follows. Section 2 describes the operational strategy of the modular bus and introduces the integrated timetabling and vehicle scheduling problem. In Section 3, an integrated model is developed to describe the joint optimization of timetable, bus formation, and vehicle scheduling in which detachment and joining at intermediate stops is allowed. To boost the computing efficiency, Section 4 introduces the ADMM decomposition algorithms. The numerical experiment results using a real-world case study are presented in Section 5. Section 6 discusses the main findings and prospective research directions.

2. Problem statement

As depicted in Fig. 5, we consider a bus loop line that comprises two opposite directions (clockwise direction and counter-clockwise direction) with their respective terminal depots and stops. Furthermore, current depots (such as those of other lines) along the loop line can serve as "intermediate depots," allowing physical modular bus units that have detached from the modular bus to be stored. We define stops located near the depots (terminal and intermediate) as "assemblage stops," where modular buses can adjust their vehicle formations. It should be noted that the term "depot(s)" in the following context refers to both terminal and intermediate depots. As a result, each direction can be divided into several segments based on pairs of adjacent assemblage stops, serviced by modular bus trips with varying vehicle formations. Under this operational scenario, modular buses with specific vehicle formations depart from the terminal depot of the direction and can adjust their vehicle formation at each assemblage stop to meet the time-dependent passenger demand on the corresponding segments before returning to the terminal depot of the direction.

This paper introduces an optimization model that integrates timetable, bus formation, and vehicle scheduling into a unified framework, depicted in Fig. 4(b). The aim is to minimize the total cost of passenger waiting time cost, operational costs, and penalty costs associated with detachment and joining operations. The model considers various constraints, including strict vehicle capacity, minimum and maximum headway, and a limited number of available modular bus units. Our study simultaneously tackles the timetable and bus formation plan subproblem as well as the modular bus unit scheduling subproblem. The primary decisions for the timetable and bus formation subproblem encompass the dispatching time and the vehicle formation at each assemblage stop for every modular bus trip. The vehicle scheduling subproblem involves assigning modular bus units to trips. Unlike traditional vehicle scheduling, where precisely one bus covers a single trip, a modular bus comprises several modular bus units. To cover a modular bus trip, modular bus units are assigned from a depot to the assemblage stop via the outgoing depot arcs to join a modular bus. The modular bus then serves all stops within a segment in a specific direction, as depicted in Fig. 5. Upon serving the segment, modular bus units are either assigned to continue with the trip on the next segment or detach from the modular bus at the assemblage stop (the segment's destination stop) according to the timetable and bus formation plan. Due to the restricted space at the assemblage stop, modular bus units that detach from the modular bus return to the depot empty via the incoming depot arcs and wait for another trip.

3. Model formulation

This section begins by introducing the assumptions underlying our model. Next, we present an STS network and a space–time network, which illustrate our timetable and bus formation subproblem and the vehicle scheduling subproblem, respectively. Finally, building upon these two networks, we propose a unified model that incorporates coupling constraints to integrate the two subproblems.

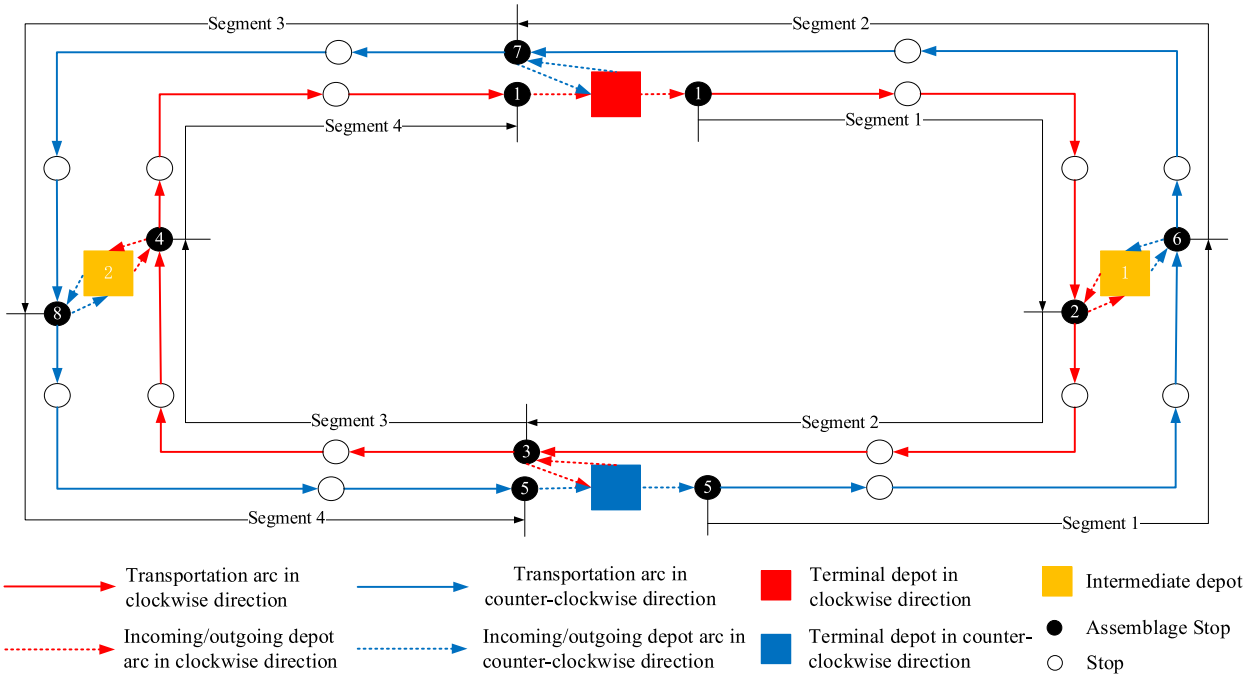


Fig. 5. Operation scenario of the considered AMV-based bus line.

3.1. Assumption

To facilitate our integrated modelling of timetable, bus formation, and vehicle scheduling, this study relies on the following five key assumptions:

1. We address a deterministic problem in which passenger demand varies over time but is known a priori, either through historical data or predicted passenger demand. This approach does not account for potential uncertainties in passenger demand, but it is a suitable method for tactical level problems that rely mainly on the expected demand and are resilient to minor deviations caused by demand uncertainty. Historical or predicted demand data are commonly used by transit operators for timetabling problems (Chen and Li, 2021; Shi and Li, 2021; Sun et al., 2014). While stochastic programming approaches can incorporate demand uncertainty, they fall outside the scope of this paper.
2. Modular buses with specific vehicle formations depart from the terminal depot and can adjust their formation at the assemblage stop near the depots. However, the availability of adequate parking space for detached modular bus units at each stop cannot be guaranteed. As a result, it is assumed that once detached from a modular bus, the modular bus units will promptly proceed to the adjacent depot affiliated with the assemblage stops. Furthermore, it is assumed that the capacity of each depot to accommodate modular bus units is unlimited, aligning with the established facts.
3. Modular bus units can detach from a modular bus and join another modular bus in either direction, while the modular bus it was previously part of continues its route without any delay.
4. The bus dwell time at each stop is constant, which is sufficient for all passengers to board and alight from the bus. This assumption holds validity, particularly for modular bus units with small capacity (Repoux et al., 2021; Tian et al., 2023). Furthermore, the travel time (including the dwell time) between any two consecutive stops is time-dependent and only depends on the dispatching time of modular buses. This assumption is reasonable as travel time is influenced by traffic volume and consequently fluctuates over time (Shi and Li, 2021). Historical bus GPS trajectory data allows for the deduction of such variations.
5. All passenger waiting at stops must be served at the end of the planning horizon (Niu and Zhou, 2013).

In our study, we present a novel modelling technique, motivated by Sun et al. (2014), to address the demand-driven timetabling problem. As depicted in Fig. 6(a), the travel time from stop 1 to stop s along the route, for a vehicle dispatched at time t, is defined as the operational offset of stop s corresponding to dispatching time t. By subtracting the operational offset of stop s, we can adjust the arrival time of vehicles 1 and 2, as well as the arrival time of passengers they will meet at stop s, as demonstrated in Fig. 6(b). This approach results in equivalent arrival time, enabling both vehicles 1 and 2 to arrive simultaneously at each stop during their respective operation processes. Consequently, we can simplify the formulation of the timetabling problem by optimizing only the dispatching time of each vehicle. Note that the operational offsets are different because we consider time-dependent travel times that are contingent upon the dispatching time of the vehicles. This arrival time revision technique has been demonstrated in previous studies by Chen and Li (2021), Shi and Li (2021), and Yang et al. (2017).

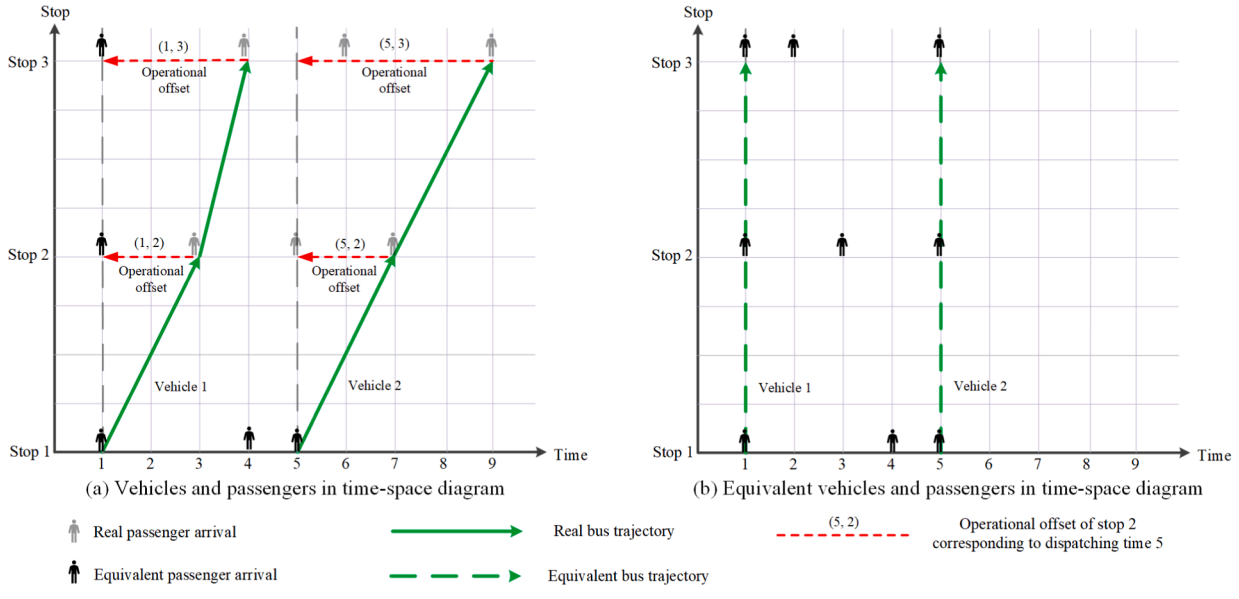
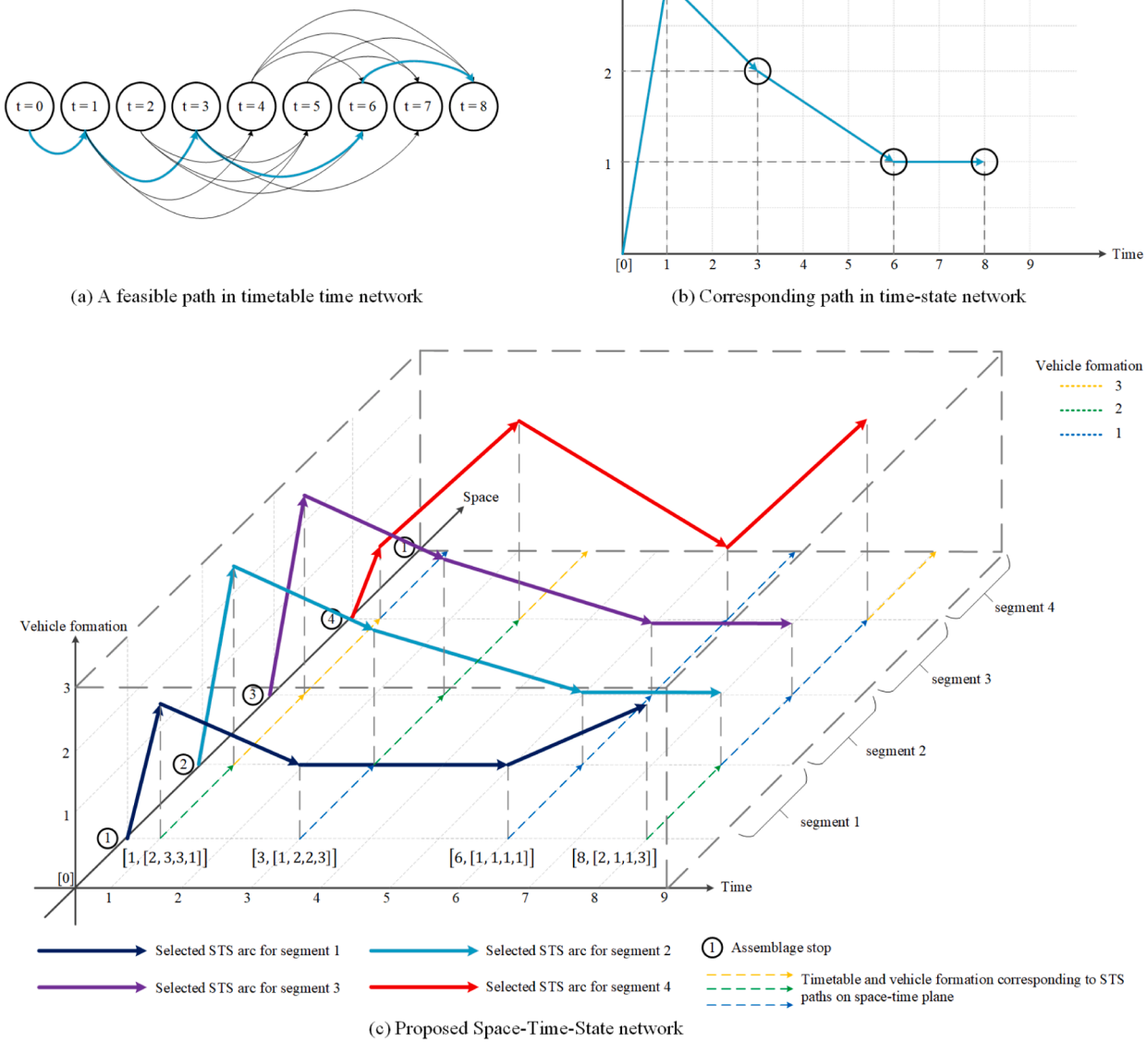


Fig. 6. The illustration diagram for equivalent arrival time of vehicles and passengers.

3.2. STS network for modular bus service

The timetabling problem of a bus route can be represented as a network flow problem through a directed graph $G = (N, A)$ (Hassold and Ceder 2012; Carosi et al. 2019). The set $N = \{t_0, 1, 2\cdots, |T|\}$ denotes the bus dispatching time nodes where t_0 and $|T|$ represent the virtual start time and the end time of the route's planning horizon, respectively. The set $A = \{(t_1, t_1^{'}), (t_2, t_2^{'}), \cdots (t_m, t_m^{'})\}$ signifies the feasible time arcs, with $(t_m, t_m^{'})$ indicating that a bus dispatched at time t_m is followed by another bus dispatched at time $t_m^{'}$, in compliance with a series of constraints such as the minimum/maximum headway (Hassold and Ceder, 2012). Thus, the timetable begins at the virtual start time (source), explores feasible bus dispatching times through directed feasible time arcs and ends at the end time (sink). The selected bus dispatching times are linked by feasible time arcs to form a path in the network. Fig. 7(a) displays a simple timetable network with a feasible path, where $h_{min} = 2$ is the minimum headway, $h_{max} = 4$ is the maximum headway, and $N = \{t_0, 1, 2, 3, 4, 5, 7, 8\}$.



(a) A feasible path in timetable time network

(b) Corresponding path in time-state network

(c) Proposed Space-Time-State network

Fig. 7. The modular bus service STS network building methodology.

This study focuses on the operational planning of a modular bus system, which require determining dispatching times and designing flexible bus formations. To address these issues, we extend the time network by adding a state dimension to represent the bus formation. This results in a time-state network, which is shown in Fig. 7(b). For each time node t, a sequence of time-state nodes (t,p) is generated, where p represents the number of modular bus units that form a modular bus, also known as bus formation. The bus formation $p \in P = \{p_{min}, p_{min}+1, \cdots, p_{max}\}$, where p_{min} and p_{max} denote the minimum and maximum allowed bus formations, respectively. Fig. 7(b) illustrates a feasible time-state path for dispatching modular buses, corresponding to the timetable depicted in Fig. 7 (a).

A time-state network is established at each stop along the route to illustrate the arrival time and vehicle formation of each modular bus. Due to arrival time revision technique, it is unnecessary to track arrival times of modular buses at each stop. However, the formation of the modular buses may vary along the route, causing the time-state paths to differ at various stops. It is important to note that intermediate stops in a segment share the same time-state path as the first stop (assemblage stop) of the segment. This is due to modular buses being unable to modify their formations at stops other than the assemblage stops. As a result, only the time-state network at each assemblage stop needs to be modeled. Each assemblage stop located at position i on a given route has a unique time-state path (represented by the solid line in Fig. 7(c)) within the time-state plane. We introduce an additional space dimension to the time-state network proposed earlier, creating an STS network that enables the intuitive depiction of changes in vehicle formation at assemblage stops. The arcs (i,i,t,t',p,p') in the STS network connect two STS nodes (i,t,p) and (i,t',p'), representing the arrival times (t and t') and vehicle formations (p and p') of two consecutive modular buses arriving at assemblage stop i.

Fig. 7(c) depicts the selected STS paths in the STS network with four assemblage stops. Each STS path corresponds to the timetable and the bus formation plan for each assemblage stop along the route. To account for the modular bus's simultaneous arrival at each stop, albeit with varying formations on different segments, we utilize a state vector $\boldsymbol{p} = [p_1, \cdots, p_i]$ that indicates the vehicle formation p_i of modular buses at each assemblage stop i. For example, consider the modular bus service depicted in Fig. 7(c) that dispatched at time 1. A modular bus is dispatched at time 1 with vehicle formation 2 from the first stop, then proceeds to pick up 1 modular bus unit at assemblage stop 2 and drop off 2 modular bus units at assemblage stop 4. The state vector representing this trip is $\boldsymbol{p} = [2,3,3,1]$. To represent a trip starting at time t, the STS nodes (i,t,p) can be rewritten in a concise form as $(t,\boldsymbol{p})$. The modular bus timetable and vehicle formation plan diagram, corresponding to the selected STS paths, can be visualized as dashed lines in the time–space plane, which provides an intuitive representation.

3.3. Time-space network for modular bus unit scheduling

In our study, modular bus units can depart individually or combine with other units from the depot to cover either the entire route or portion of it. After finishing a trip, modular bus units return to the depot from the trip's destination stop, where they wait for their next trips. To address this multi-depot bus scheduling problem, we utilize a time–space network modeling approach, as previously demonstrated in the literature (Kliewer et al., 2006; Liu et al., 2023). In order to track the modular bus units' trajectories for scheduling within the planning horizon, we expand the physical network to a time–space network by incorporating a time dimension. This expansion involves extending a physical node i to a time–space node (i,t), and a physical arc (i,j) to a time–space arc (i,j,t,t'). The time–space arc (i,j,t,t') represents the flow of vehicles departing from node i at time t and arriving at node j at time t'.

We present Fig. 8 as an example to illustrate the construction of a time–space network. In the physical transportation network (see Fig. 8(a)), a physical arc (i,j) can represent a transportation segment, an incoming/outgoing depot arc (when $i \neq j$), or a depot waiting arc (when $i = j$). As a result, modular bus units use different types of arcs during their trips, naturally forming vehicle flows in the time–space network (Fig. 8(b)). Specifically, the following arc types are defined:

- Outgoing depot arc (i, j, t, t'): represents the process of modular bus units departing from depot i at time t and traveling to their first stop (assemblage stop) j on the trip, with a travel time of $(t' - t)$.
- Transportation arc (i, j, t, t'): represents the traveling process of modular bus units traversing a segment between two adjacent assemblage stops, departing from assemblage stop i at time t and arriving at assemblage stop j at time t'. The time duration $(t' - t)$ of arc (i, j, t, t') is the travel time of modular bus units in the segment.
- Incoming depot arc (i, j, t, t'): represents the process of modular bus units departing from the final stop i on the trip at time t and returning to depot j, with a travel time of $(t' - t)$.
- Depot waiting arc $(i,i,t,t+1)$: represents the waiting process of modular bus units at depot i for one unit of time to prepare for their next trip. If modular bus units wait at the depot for several time units, several successive depot waiting arcs are used to denote the waiting process.
- Deadhead arc (i, j, t, t'): represents the traveling process of modular bus units traversing a segment between two depots, departing from depot i at time t and arriving at depot j at time t'. The time duration $(t' - t)$ of arc (i, j, t, t') is the travel time of modular bus units in the segment.

As illustrated in Fig. 8(b), at the beginning of the planning horizon, terminal depots 1 and 2 are equipped with eight modular bus units that are available for scheduling. These modular bus units travel across the time–space network to execute modular bus trips, which are comprised of entire and partial route trips between assemblage stops. For instance, the first upward trip involves two entire route trips (from assemblage stop 1 to assemblage stop 4) and one partial route trip (from assemblage stop 1 to assemblage stop 3) for

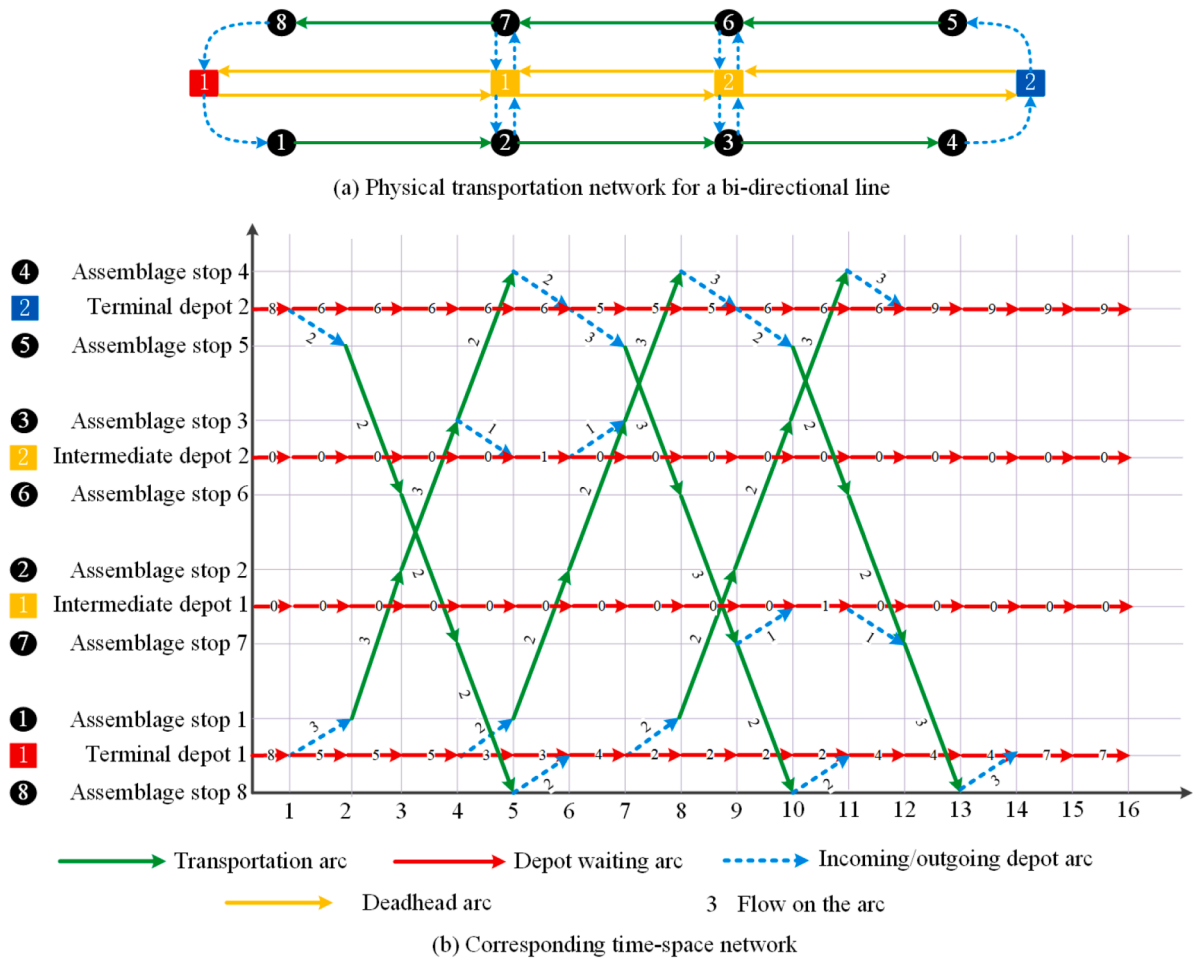


Fig. 8. The vehicle scheduling time–space network for modular bus units.

the modular bus units. As a result, three modular bus units combine to form a modular bus and depart from terminal depot 1 at time 1. Upon completion of the partial trip, one modular bus unit separates from the modular bus and returns to intermediate depot 2, while the modular bus continues along its route without delay. In this time–space network, vehicles can only move forward in time (i.e., $t^{'} > t$), and flow conservation must be ensured to maintain the modular bus unit scheduling network's closure.

3.4. Mathematical model

3.4.1. Notation

Table 1 and Table 2 summarize the sets, indices, and parameters used in this study, providing a convenient reference for readers.

3.4.2. Variables

1) Decision variables.

Based on the STS network introduced in Section 3.2 and the time–space network described in Section 3.3, and utilizing the notation specified in Table 1 and Table 2, we constructed the following binary variable denoted as $y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}$. This variable represents whether the STS arc $(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}) \in A_d$ occurs on the unique STS path that indicates the timetable and bus formation plan for direction d.

$$y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}} = \begin{cases} 1, if STS arc(t,t^{'},p,p^{'}) \in A_d occurs on the unique STS path of direction d \\ 0, otherwise \end{cases} \tag{1}$$

In order to model the flow of vehicles in the modular bus unit scheduling time–space network, we introduce a novel variable, denoted as $x^{k}_{i,j,t,t^{'}}$, which is a non-negative integer variable representing the vehicle flow originating from depot k on the time–space arc

Table 1
Notation of sets and indices in the model.

Notation	Definition
Sets:	
D	Set of directions
N	Set of physical nodes, including stops and depot
T	Set of discrete time nodes of the planning horizon
$N^{'}$	Set of time–space nodes in the time–space network of modular bus unit scheduling
P	Set of vehicle formations
P_d	Set of vehicle formation state vectors for direction d, $d \in D$
K	Set of terminal depots
$A_{sts}(d)$	Set of STS arcs in the modular bus service STS network of direction d, $d \in D$
A_{st}	Set of time–space arcs in the time–space network of modular bus unit scheduling
$A_{st}(k)$	Set of time–space arcs in the time–space network for modular bus units that belong to terminal depot k, $A_{st}(k) \subset A_{st}$
Ψ	Set of time–space transportation arcs, $\Psi \subset A_{st}$
$S(d)$	Set of stops in direction d, $S(d) \subset N$
$S^*(d)$	Set of assemblage stops in direction d, $S^*(d) = \{1, 2, \cdots \lvert S^*(d) \rvert\}$, $S^*(d) \subset S(d)$
Indices:	
d	Index of directions, $d \in D$
$i, i^{'}, j$	Index of physical nodes, $i, i^{'}, j \in N$
s	Index of stops in direction d, $s \in S(d)$
s^*	Index of assemblage stops in direction d, $s^* \in S^*(d)$
$t, t^{'}, t^{''}$	Index of time nodes, $t, t^{'}, t^{''} \in T$
p_i	Index of vehicle formation at assemblage stop i, $i \in S^*(d)$, $p_i \in P$
$\boldsymbol{p}, \boldsymbol{p}^{'}, \boldsymbol{p}^{''}$	Index of vehicle formation state vectors where $\boldsymbol{p} = \left[p_1, \cdots, p_i, \cdots, p_{\lvert S^*(d) \rvert}\right]$ indicates the modular bus's formation p_i at each assemblage stop i in direction d. Note: $i \in S^*(d)$, $p_i \in P$, $\boldsymbol{p}, \boldsymbol{p}^{'}, \boldsymbol{p}^{''} \in P_d$
k	Index of terminal depot, $k \in K$
$(t, \boldsymbol{p})$, $(t^{'}, \boldsymbol{p}^{'})$	Index of STS nodes in the modular bus service STS network, means a modular bus with vehicle formation state vector $\boldsymbol{p}$ dispatched at time t
$(t, t^{'}, \boldsymbol{p}, \boldsymbol{p}^{'})$	Index of STS arcs, means a modular bus with vehicle formation state vector $\boldsymbol{p}$ that is dispatched at time t, followed by another modular bus with vehicle formation state vector $\boldsymbol{p}^{'}$ that is dispatched at time $t^{'}$, $(t, t^{'}, \boldsymbol{p}, \boldsymbol{p}^{'}) \in A_d$
$(i, t), (j, t^{'})$	Index of time–space nodes in the modular bus unit scheduling time–space network, $(i, t), (j, t^{'}) \in N^{'}$
$(i, j, t, t^{'})$	Index of time–space arcs, means modular bus units travel from node i at time t to node j at time $t^{'}$, $(i, j, t, t^{'}) \in A_{st}$

Table 2
Notation of parameters in the model.

Notation	Definition
t_0	Virtual start time of planning horizon
$\lvert T \rvert$	End time of planning horizon
$\boldsymbol{p}_0$	Virtual vehicle formation state vector corresponding to the virtual start time
Cap	Capacity of a single modular bus unit
$\alpha_t(s)$	Number of arriving passengers at stop s at time t
$\alpha_t(s^{'}, s)$	Number of arriving passengers at stop $s^{'}$ at time t with the destination of stop s
h_{min}	Minimum dispatching headway
h_{max}	Maximum dispatching headway
p_{min}	Minimum vehicle formation
p_{max}	Maximum vehicle formation
$\tau_s(t)$	Operational offset of stop s corresponding to dispatching time t
dep_k	Earliest departure time of modular bus units at the origin terminal depot k
arr_k	Latest arrival time of modular bus units at the destination terminal depot k
v_k	Initial available number of modular bus units at depot k
c^p	Operational costs for a modular bus with the vehicle formation p
c_{jd}	Penalty cost associated with a detachment or joining operation
c_w	Unit-waiting cost per passenger per minute

$(i, j, t, t^{'}) \in A_{st}$. The value range of this variable is defined by Equation (2) and (3).

$$0 \le x^k_{i,j,t,t^{'}} \le p_{max} \forall k \in K, \forall (i, j, t, t^{'}) \in \Psi \tag{2}$$

$$x^k_{i,j,t,t^{'}} \ge 0 \forall k \in K, \forall (i, j, t, t^{'}) \in A_{st} \backslash \Psi \tag{3}$$

2) Dependent variables.

Table 3 provides a summary of the dependent variables that capture the passengers' dynamic evolution with respect to the

Table 3
Notation of dependent variables in the model.

Notation	Definition
$W_t(s)$	Number of passengers waiting for the modular bus at stop s at time t
$F_t(s)$	Number of passengers failing to board the modular bus at stop s at time t
$B_t(s)$	Number of passengers who can board the modular bus at stop s at time t
$A_t(s)$	Number of passengers alighting from the modular bus at stop s at time t
$I_t(s)$	Number of passengers on the modular bus departing from stop s at time t
$\beta_t(s^{'},s)$	Ratio of passengers boarding the modular bus at stop $s^{'}$ at time t and travelling to stop s

timetable and bus formation plan variables $y^d_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}$.

3.4.3. Constraints on the timetable and bus formation plan

(1) Modular bus dispatching flow conservation constraints.

In Section 3.2, we have presented the STS network as a fundamental component of our approach to optimize timetable and bus formation plan. To achieve this, we employ a set of standard flow conservation constraints (4)-(6) for each direction $d \in D$, enabling us to identify a unique path in the STS network and generate a corresponding timetable and bus formation plan. Constraint (4) indicates that the direction d initiates the dispatching of modular buses after the virtual start time t_0. Constraint (5) ensures that the inflow of each intermediate node in the STS network equals the outflow, thereby preserving the continuity and uniqueness of the derived timetable and bus formation plan. Lastly, Constraint (6) ensures that the timetable concludes at the specified time $|T|$, and that all passengers can be served within the planning horizon.

$$\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)} y^d_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}} = 1t = t_0, \boldsymbol{p} = \boldsymbol{p}_0, \forall d \in D \tag{4}$$

$$\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)} y^d_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}} - \sum_{(t^{*},t,\boldsymbol{p}^{'},\boldsymbol{p})\in A_{sts}(d)} y^d_{t^{'},t,\boldsymbol{p}^{'},\boldsymbol{p}} = 0 \forall t \notin \{t_0, |T|\}, \forall \boldsymbol{p} \in P_d, \forall d \in D \tag{5}$$

$$\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)} y^d_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}} = 1t^{'} = |T|, \forall d \in D \tag{6}$$

(2) Passengers' dynamic evolution constraints.

In oversaturated conditions, it is a common occurrence that passengers are unable to board the first arriving bus they meet at a stop, resulting in extended waiting times. Thus, accurately characterizing and modeling the dynamic changes in passenger volume at each stop is crucial. Specifically, for an STS arc $(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}) \in A_d$, the number of passengers waiting for a modular bus at time $t^{'}$ at stop s is denoted by $W_{t^{'}}(s)$. This value is calculated by summing the failing-to-board passengers from the previous modular bus that was dispatched at time t (denoted by $F_t(s)$) and the new arriving passengers within the time interval $[t,t^{'}]$, i.e., $\sum_{tt=t}^{t^{'}}\alpha_{tt}(s)$, where $\alpha_{tt}(s)$ represents the number of arriving passengers at stop s at time tt. For the first modular bus that was dispatched at time $t^{'}$, $W_{t^{'}}(s)$ only includes the number of arriving passengers at each stop. The total number of passengers waiting for a modular bus at time $t^{'}$ at stop s on the STS arc $(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})$ can be expressed using the following Equation (7).

$$W_{t^{'}}(s) = \begin{cases} \sum_{tt=0}^{t^{'}}\alpha_{tt}(s), if t = t_0 \\ F_t(s) + \sum_{tt=t}^{t^{'}}\alpha_{tt}(s), otherwise \end{cases}, \forall d \in D, \forall s \in S(d) \tag{7}$$

Besides, for an STS arc $(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}) \in A_d$, $\beta_t(s^{'},s)$ represents the ratio of passengers boarding the modular bus at stop $s^{'}$ at time $t^{'}$ and alighting at stop s. This value is computed as the number of passengers at stop $s^{'}$ at time $t^{'}$ who are alighting at stop s, divided by the total number of passengers at stop $s^{'}$ at time $t^{'}$ (denoted by $W_{t^{'}}(s^{'})$). Additionally, for the first modular bus that was dispatched at time $t^{'}$, the calculation of $\beta_{t^{'}}(s^{'},s)$ takes into account only the new arriving passengers within the time interval $[t,t^{'}]$, denoted by $\sum_{tt=t}^{t^{'}}\alpha_{tt}(s^{'},s)$, where $\alpha_{tt}(s^{'},s)$ represents the number of arriving passengers at stop $s^{'}$ at time tt with the destination of stop s. As for other time, the calculation of $\beta_{t^{'}}(s^{'},s)$ not only take into account the new arriving passengers within the time interval $[t,t^{'}]$ but also the failing-to-board passengers from the previous modular bus that was dispatched at time t (denoted by $F_t(s)$). Mathematically, $\beta_{t^{'}}(s^{'},s)$ can be expressed using the following Equation (8).

$$\beta_t(s',s)=\begin{cases}\dfrac{\sum_{tt=0}^{t}\alpha_{tt}(s',s)}{W_t(s')}, if\ t=t_0\\ \dfrac{\sum_{tt=t}^{t}\alpha_{tt}(s',s)+\beta_t(s',s)\bullet F_t(s)}{W_t(s')}, otherwise\end{cases},\forall d\in D,\forall s,s'\in S(d):s'<s \tag{8}$$

The number of passengers at stop s who failed to board the previous modular bus that was dispatched at time t is denoted by $F_t(s)$. To provide a more intuitive representation, we employ the cumulative passenger arrival curves and cumulative passenger departure curves illustrated in Fig. 9 to exemplify the computation of $F_t(s)$. This quantity is equal to the difference between the number of passengers waiting for the bus (denoted by $W_t(s)$) and the number of passengers who were able to board the bus (denoted by $B_t(s)$). The calculation of $F_t(s)$ can be obtained using the following Equation (9).

$$F_t(s)=W_t(s)-B_t(s),\forall d\in D,\forall s\in S(d) \tag{9}$$

Here, $W_t(s)$ represents the number of passengers at stop s waiting for the modular bus that was dispatched at time t. It can be calculated by Equation (7).

At stop s, $B_t(s)$ represents the number of passengers who can board the modular bus dispatched at time t. This value is determined by the number of waiting passengers and the remaining capacity of the arriving modular bus with vehicle formation $p_t(s)$. Specifically, if the number of waiting passengers is less than the remaining capacity, all waiting passengers will board the modular bus. However, if the number of waiting passengers exceeds the remaining capacity, only a portion of them will be able to board, while the remaining passengers must wait for the next bus. In other words, the number of boarding passengers is given by the minimum value between the number of waiting passengers and the remaining capacity, which can be expressed as follows.

$$B_t(s)=min(W_t(s),cap\bullet p_t(s)+A_t(s)-I_t(s-1)),\forall d\in D,\forall s\in S(d) \tag{10}$$

For the modular bus dispatched at time t, the number of alighting passengers at stop s (denoted by $A_t(s)$) is the sum of passengers who boarded this modular bus at upstream stops and are heading towards stop s, which can be expressed as follows.

$$A_t(s)=\sum_{s'\in S(d):s'<s}[B_t(s')\bullet\beta_t(s',s)],\forall d\in D,\forall s\in S(d) \tag{11}$$

Where, $\beta_t(s', s)$ indicates the ratio of passengers boarding this modular bus at stop s' with the destination of stop s.

Moreover, while a modular bus is dwelling at a stop, passengers may need to alight from the vehicle if the current stop is their destination. At the same time, waiting passengers at the stop can board the modular bus if there is available capacity. Therefore, the number of in-vehicle passengers when the modular bus departs from stop s at time t varies and can be defined as follows.

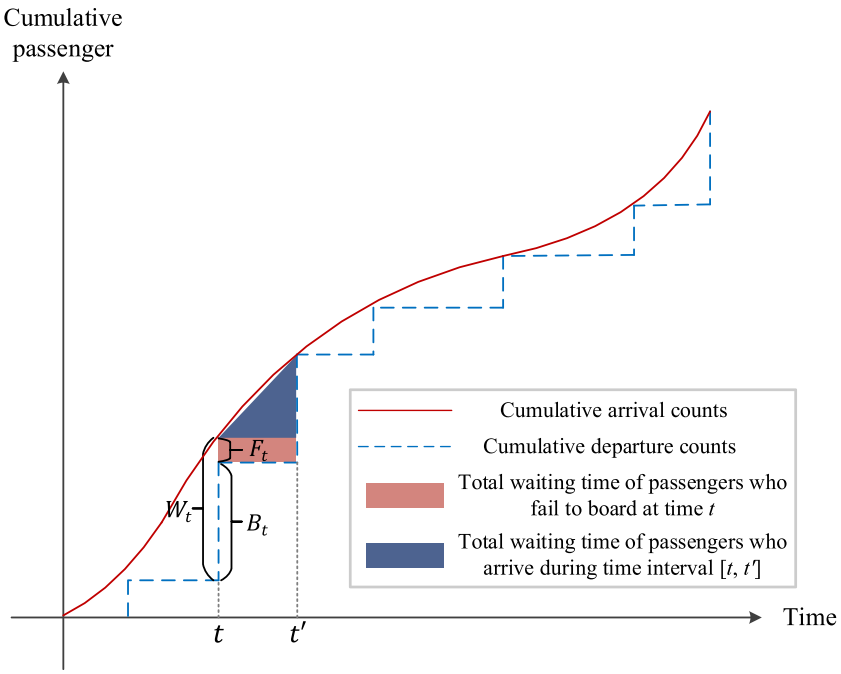


Fig. 9. Dynamic evolution of passenger arrivals and departures at a stop.

$$I_t(s) = \begin{cases} B_t(s), if\ s = 1 \\ I_t(s-1) - A_t(s) + B_t(s), otherwise \end{cases}, \forall d \in D, \forall s \in S(d) \tag{12}$$

(3) Dispatching headway constraints.

In the modular bus service STS network, the dispatching headway h for a given arc $(t, t', \boldsymbol{p}, \boldsymbol{p}')$ is calculated as $t' - t$. To ensure the safety and efficiency for the modular bus service, it is important to establish minimum and maximum dispatching headway constraints. These constraints can be enforced through Rule 1 as follows:

Rule 1: For each STS arc $(t, t', \boldsymbol{p}, \boldsymbol{p}')$, the dispatching headway h should not be smaller than the minimum headway, h_{min}, and no larger than the maximum headway h_{max}, that is, $h_{min} \le h \le h_{max}$. Any arc that fails to meet this constraint will be removed from the network.

In summary, embedding the minimum and maximum dispatching headway directly into the solution method in the form of Rule 1 eliminates infeasible STS arcs during the network construction phase, reducing the search space and improving computation efficiency.

3.4.4. Constraints on the modular bus unit scheduling problem

The scheduling of modular bus units is formulated in the time–space network presented in Section 3.3. The flow of modular bus units on arc (i, j, t, t') originating from depot k is represented by $x^k_{i,j,t,t'}$, subject to Constraints (13) and (14). The parameter v_k denotes the initial available number of modular bus units at depot k. Constraint (13) ensures that the number of modular bus units used for trips from each depot k does not exceed v_k, where dep_k is the earliest departure time. Constraint (14) is the flow conservation constraint for the intermediate node, where arr_k is the latest arrival time. The flow on each time–space arc corresponds to multiple unique time–space paths (trip chains) of individual modular bus units originating from each depot in the time–space network.

$$\sum_{(i,j,t,t') \in A_{st}(k)} x^k_{i,j,t,t'} \le v_k i = k, t = dep_k, \forall k \in K \tag{13}$$

$$\sum_{(i,j,t,t') \in A_{st}(k)} x^k_{i,j,t,t'} - \sum_{(i',i,t'',t) \in A_{st}(k)} x^k_{i',i,t'',t} = 0 \forall (i,t) \notin \{(k, dep_k), (k, arr_k)\}, \forall k \in K \tag{14}$$

3.4.5. Vehicle resources consistency constraint

It is crucial to assign the exact number of modular bus units required for each modular bus service in a feasible timetable and bus formation plan for each direction. To achieve this, Constraint (15) ensures that the decision variables of modular bus unit scheduling are consistent with the timetable and bus formation plan. Additionally, according to the modular bus dispatching flow conservation Constraints (4) – (6), Constraint (15) also makes sure that only one modular bus can be dispatched at any given time (the right-hand side of the equation is not greater than 1).

Specifically, if $y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'} = 1$, it indicates that a modular bus is dispatched at time t with vehicle formation state vector $\boldsymbol{p}$ in direction d. Here, $\boldsymbol{p} = \left[p_1, \cdots, p_{s^*}, \cdots, p_{|S^*(d)|}\right]$ $(\forall s^* \in S^*(d))$ represents the flexible formation of the modular bus on different segments. Constraint (15) enforces that p_{s^*} modular bus units must arrive at the assemblage stop s^* at time $t + \tau_{s^*}(t)$ to form the modular bus with specific vehicle formation p_{s^*}, where $\tau_{s^*}(t)$ is the operational offset of stop s^* corresponding to dispatching time t. In other words, the vehicle flow on the transportation arc originating from the assemblage stop s^* in direction d (denoted by $\Psi^d_{s^*}$) at time $t + \tau_{s^*}$ must be p_{s^*}; Otherwise, if $\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}') \in A_{sts}(d)} (p_{s^*} \times y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}) = 0$, no modular buses are dispatched at time t, and modular bus units cannot travel on the transportation arc belonging to $\Psi^d_{s^*}$ at time $t + \tau_{s^*}(t)$.

$$\sum_{(i,j,t'',t') \in \Psi^d_{s^*}, t' = t + \tau_{s^*}(t)} \sum_{k \in K} x^k_{i,j,t'',t'} = \sum_{(t,t',\boldsymbol{p},\boldsymbol{p}') \in A_{sts}(d)} \left(p_{s^*} \bullet y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right) \forall d \in D, \forall s^* \in S^*(d), \forall t \in T \tag{15}$$

3.4.6. Objective function

Here, the objective function of the mathematical model for the integrated optimization of timetable, bus formation, and vehicle scheduling are introduced in detail.

(1) Passenger waiting time cost.

To determine the cost of passenger waiting time, it is necessary to calculate the duration of time that passengers spend waiting for the modular bus to arrive. Specifically, the waiting passengers on an STS arc $(t, t', \boldsymbol{p}, \boldsymbol{p}') \in A_d$ are comprised of two parts: those who were

unable to board the previous modular bus dispatched at time t, and those who arrive during the time interval $[t, t']$. Therefore, the passenger waiting time can be divided into two parts. The first part is calculated as the number of passengers who failed to board the previous bus, multiplied by the headway duration, i.e., $(t' - t) \bullet F_t(s)$. The second part of passenger waiting time can be approximated as half of the headway duration multiplied by the total number of waiting passengers, i.e., $0.5 \cdot (t' - t) \bullet \sum_{tt=t}^{t'} \alpha_{tt}(s)$. This approach has been adopted in previous studies as an effective way to reduce computation time for high-frequency transit when calculating passenger waiting time (Chen and Li, 2021; Tian et al., 2023). By combining these two parts, we can formulate Equation (16) for the total cost of passenger waiting time, where c_w represents the unit-waiting cost per passenger per minute.

$$Z_1 = c_w \bullet \sum_{d \in D} \sum_{(t,t',\boldsymbol{p},\boldsymbol{p}') \in A_{sts}(d)} \left[y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'} \bullet (t' - t) \bullet \sum_{s \in S(d)} \left(F_t(s) + 0.5 \sum_{tt=t}^{t'} \alpha_{tt}(s) \right) \right] \tag{16}$$

(2) Vehicle operational costs.

In contrast to a fixed-capacity bus, a modular bus with varying formations incurs significantly different operational costs. As established by Chen et al. (2019), the operational costs c_p of a modular bus with a vehicle formation comprising p modular bus units can be calculated using Equation (17).

$$c^p = C^F + C^V \bullet (p \bullet Cap)^{\alpha} \forall p \in P \tag{17}$$

Here, C^F denotes the fixed operational cost, while C^V represents the marginal operational cost for a modular bus. The parameter $\alpha (0 < \alpha \leq 1)$ ensures that the operational costs of a modular bus exhibit concavity with respect to its capacity, thus capturing the economies of scale associated with its flexible formations (Chen and Li, 2021; Dai et al., 2020). In this study, we employ a linear function ($\alpha = 1$) to calculate operational costs.

It is noteworthy that a modular bus can change its vehicle formation at assemblage stops during a single service. As a result, the total operational costs for a modular bus over the entire route in direction d can be calculated using Equation (18).

$$c^{\boldsymbol{p}}_d = c^{p_1} + \cdots + c^{p_i} + \cdots + c^{p_{|S^*(d)|}} \forall d \in D, \forall i \in S^*(d), \forall p_i \in P \tag{18}$$

Here, the formation of the modular bus at each assemblage stop i during a service in direction d is denoted by $\boldsymbol{p} = \left[p_1, \cdots, p_i, \cdots, p_{|S^*(d)|} \right]$.

(3) Penalty costs associated with detachment and joining operations.

The en-route physical detachment and joining of AMVs may pose a certain level of risk, and the in-motion transfer could potentially cause discomfort or inconvenience for some passengers. To mitigate these issues, a penalty cost for each detachment or joining operation is incorporated into the objective function to avoid them. In our study, modular buses can change vehicle formation multiple times along the route during a single service. Thus, we need to track every modular bus to determine whether it changes its vehicle formation at each assemblage stop. To compute the number of detachment and joining operations $\sigma^{\boldsymbol{p}}_d$ for a modular bus service is a simple process in our proposed STS network. This is achieved by utilizing the vehicle formation state vector $\boldsymbol{p}$ of a modular bus service in direction d, where $\boldsymbol{p} \in P_d$. For instance, the STS node $(1, [2, 3, 3, 1])$ in Fig. 7(c) indicates a modular bus service dispatched at time 1 with the vehicle formation state vector $[2, 3, 3, 1]$. The number of detachment and joining operations $\sigma^{[2,3,3,1]}_d$ is 2, including one joining operation at the second assemblage stop, and one detachment operation at the fourth assemblage stop. The total penalty costs of detachment and joining operations for this service is $2 \times c_{jd}$, where c_{jd} represents the penalty cost for a detachment or joining operation.

To summarize, we present the overall objective function of our mathematical model in Equation (19). The objective function comprises of three components: the total cost of passenger waiting time, vehicle operational costs, and penalty costs incurred due to detachment and joining operations.

$$\min Z = \sum_{d \in D} \sum_{(t,t',\boldsymbol{p},\boldsymbol{p}') \in A_{sts}(d)} y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'} \left[c_w \bullet (t' - t) \bullet \sum_{s \in S(d)} \left(F_t(s) + 0.5 \sum_{tt=t}^{t'} \alpha_{tt}(s) \right) + c^{\boldsymbol{p}}_d + c_{jd} \bullet \delta^{\boldsymbol{p}}_d \right] \tag{19}$$

4. Solution approach

Our proposed model integrates the timetable and bus formation plan problem and vehicle scheduling problem by means of a set of "hard" equality constraints that couple variables $\boldsymbol{x}$ and $\boldsymbol{y}$, making it difficult to solve in large-scale cases. ADMM is a dual decomposition

method that combines Augmented Lagrangian relaxation with the block coordinate descent method (Bertsekas, 1997). It has the same form as Augmented Lagrangian relaxation but can update variables sequentially in a block-by-block scheme, updating variables in a block at each iteration with variables in other blocks fixed (Boyd et al., 2011; Yao et al., 2019). This rolling update scheme has been applied to solve the integrated planning problem with subproblems at different stages, such as train rescheduling and dynamic passenger assignment (Zhan et al., 2021), as well as line planning and train timetabling (Zhang et al., 2022). The timetable and bus formation plan problem, as well as the vehicle scheduling problem, are inherently hierarchical in nature but necessitate joint optimization. We present an ADMM-based dual decomposition algorithm to relax the vehicle resources consistency Constraint (15) and decompose our model into two subproblems, which are easier to solve than the primal model. We develop an iterative solution framework to update the two subproblems in a rolling fashion. While Chen et al. (2016) have demonstrated that ADMM is not always guaranteed to converge for convex minimization problems with more than two blocks, ADMM has proven to be a useful approach for addressing various practical problems to achieve a better upper bound solution of the primal problem within a specified maximum number of iterations. Examples of such problems include vehicle routing problems (Yao et al., 2019), yard crane and automated guided vehicle scheduling (Chen et al., 2020a), and timetabling problems (Zhang et al., 2019). We then apply the Lagrangian relaxation (LR) algorithm to decompose the primal model to obtain a lower bound solution, which is used to determine the effectiveness of the solution obtained from the ADMM method.

4.1. ADMM-based solution framework

In principle, ADMM is applied to handle equality constraints. We can directly introduce two parameters for Constraint (15), the Lagrangian multiplier $\lambda_{d,s^*,t}$ and the quadratic penalty parameter ρ, which allows the constraints to be relaxed, augmented, and incorporated into the objective function. To clearly present the solution approach, we denote the arc $(i,j,t,t^{'})$ in the vehicle scheduling time–space network as a for simplicity. Additionally, we denote the set of transportation arc originating from assemblage stop s^* at time $t+\tau_{s^*}(t)$ in direction d as $\Psi^{d}_{s^*,t}$.

The Augmented Lagrangian Relaxation (ALR) model is as follows:

$$minZ = \sum_{d\in D}\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)} y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\left[c_w \bullet (t^{'}-t)\bullet \sum_{s\in S(d)}\left(F_t(s)+0.5\sum_{tt=t}^{t^{'}}\alpha_{tt}(s)\right)+c^{\boldsymbol{p}}_{d}+c_{jd}\bullet\delta^{\boldsymbol{p}}_{d}\right]+\sum_{d\in D}\sum_{s^*\in S^*(d)}\sum_{t\in T}\lambda_{d,s^*,t}\left[\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_d}\left(p_{s^*}\bullet y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\right)-\sum_{a\in\Psi^{d}_{s^*,t}}\sum_{k\in K}x^{k}_{a}\right]+0.5\rho\sum_{d\in D}\sum_{s^*\in S^*(d)}\sum_{t\in T}\left\|\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_d}\left(p_{s^*}\bullet y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\right)-\sum_{a\in\Psi^{d}_{s^*,t}}\sum_{k\in K}x^{k}_{a}\right\|^{2}_{2} \tag{20}$$

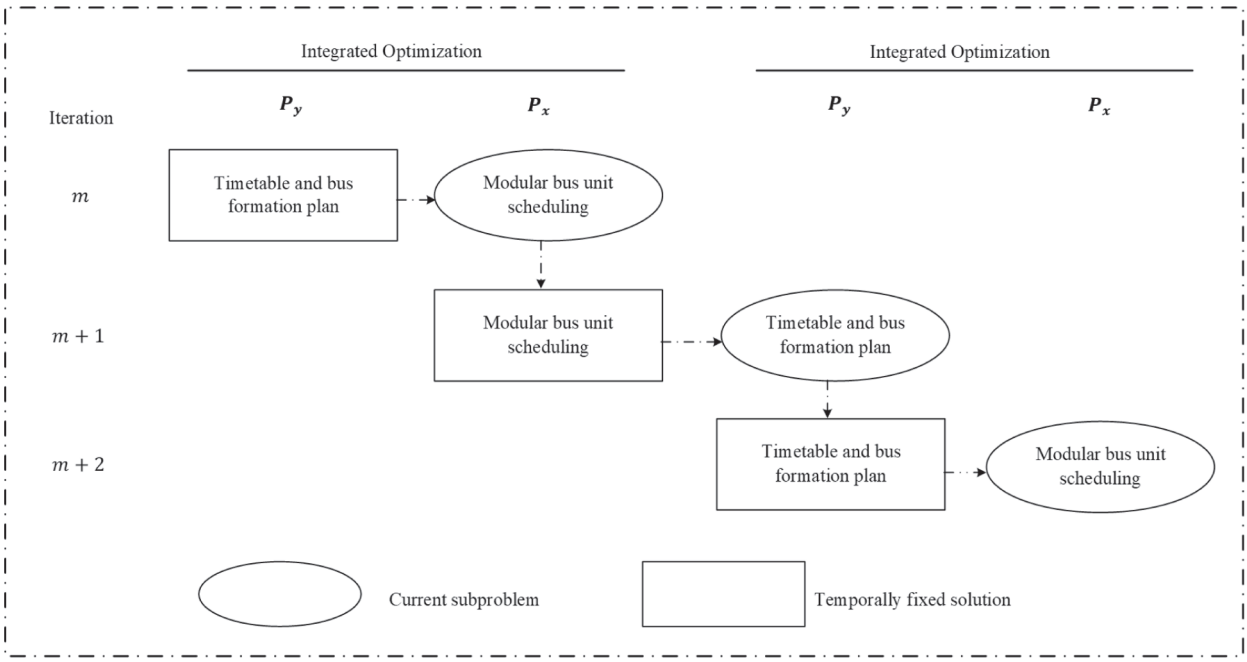


Fig. 10. The framework for the iterative solution of the two subproblems.

Subject to Constraints (4)-(14).

The ALR model includes two types of variables, $\boldsymbol{x}$ and $\boldsymbol{y}$. Using the ADMM dual decomposition method, the ALR model can be decomposed into two subproblems: the timetable and bus formation plan subproblem ($\boldsymbol{P_y}$) and the modular bus unit scheduling subproblem ($\boldsymbol{P_x}$). An iterative block coordinate descent approach is employed to solve these two subproblems. As shown in Fig. 10, when the modular bus unit scheduling subproblem $\boldsymbol{P_x}$ is solved in iteration m, the variables y for the timetable and bus formation plans subproblem $\boldsymbol{P_y}$ are temporarily fixed in the current iteration. After solving $\boldsymbol{P_x}$ in iteration m, the solution of $\boldsymbol{P_y}$ is obtained in iteration $m+1$ with variables $\boldsymbol{x}$ fixed to their values obtained in iteration m. Once $\boldsymbol{P_y}$ is solved in iteration $m+1$, its solution is temporarily fixed and $\boldsymbol{P_x}$ is solved in iteration $m+2$. We denote the fixed value of variables $\times$ and y as $\overline{x}^k_{i,j,t,t'}$ and $\overline{y}^d_{t,t,\boldsymbol{p},\boldsymbol{p}'}$, respectively.

4.1.1. Timetable and bus formation plan subproblem

The sub-model for the timetable and bus formation plan subproblem $\boldsymbol{P_y}$ in our iterative approach is presented as follows:

$$\begin{aligned} min\boldsymbol{P_y} = &\sum_{d\in D}\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_{sts}(d)} y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\left[c_w \bullet (t'-t)\bullet\sum_{s\in S(d)}\left(F_t(s)+0.5\sum_{tt=t}^{t'}\alpha_{tt}(s)\right)+c^{\boldsymbol{p}}_d+c_{jd}\bullet\delta^{\boldsymbol{p}}_d\right]+\sum_{d\in D}\sum_{s^*\in S^*(d)}\sum_{t\in T}\lambda_{d,s^*,t}\left[\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_d}\left(p_{s^*}\right.\right.\\ &\left.\left.\bullet\, y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right)-\sum_{a\in\Psi^d_{s^*,t}}\sum_{k\in K}\overline{x}^k_a\right]+0.5\rho\sum_{d\in D}\sum_{s^*\in S^*(d)}\sum_{t\in T}\Big\|\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_d}\left(p_{s^*}\bullet y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right)-\sum_{a\in\Psi^d_{s^*,t}}\sum_{k\in K}\overline{x}^k_a\Big\|_2^2 \end{aligned} \tag{21}$$

Subject to Constraints (4)-(12).

Since subproblem $\boldsymbol{P_y}$ consists solely of flow conservation constraints and passengers' dynamic evolution constraints for each direction d, it can be decomposed further into subproblems specific to individual directions, as follows:

$$\begin{aligned} min\boldsymbol{P_y}(d) = &\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_{sts}(d)} y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\left[c_w \bullet (t'-t)\bullet\sum_{s\in S(d)}\left(F_t(s)+0.5\sum_{tt=t}^{t'}\alpha_{tt}(s)\right)+c^{\boldsymbol{p}}_d+c_{jd}\bullet\delta^{\boldsymbol{p}}_d\right]+\sum_{s^*\in S^*(d)}\sum_{t\in T}\lambda_{d,s^*,t}\left[\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_d}\left(p_{s^*}\right.\right.\\ &\left.\left.\bullet\, y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right)-\sum_{a\in\Psi^d_{s^*,t}}\sum_{k\in K}\overline{x}^k_a\right]+0.5\rho\sum_{s^*\in S^*(d)}\sum_{t\in T}\Big\|\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_d}\left(p_{s^*}\bullet y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right)-\sum_{a\in\Psi^d_{s^*,t}}\sum_{k\in K}\overline{x}^k_a\Big\|_2^2 \end{aligned} \tag{22}$$

Subject to Constraints (4)-(12) for each direction d.

Neglecting the quadratic terms in the objective function, $\boldsymbol{P_y}(d)$ represents a generalized time-dependent shortest path problem in an STS network for direction d. The problem can be efficiently solved using a DP algorithm. Since all the variables in $\boldsymbol{P_y}(d)$ are binary, the quadratic terms in Equation (22) can be transformed into linear terms. This conversion can be achieved by introducing a constant term, denoted by $\overline{R}^d_{s^*,t}$, which denotes the flow of modular bus units on the transportation time–space arc $a\in\Psi^d_{s^*,t}$. For the sake of simplicity, $\sum_{a\in\Psi^d_{s^*,t}}\sum_{k\in K}\overline{x}^k_a$ is replaced with $\overline{R}^d_{s^*,t}$.

$$\begin{aligned} \Big\|\sum_{(t,t',p,p')\in A_{sts}(d)}\left(p_{s^*}\cdot y^d_{t,t',p,p'}\right)-\overline{R}^d_{s^*,t}\Big\|_2^2 &= \left[\sum_{(t,t',p,p')\in A_{sts}(d)}\left(p_{s^*}\cdot y^d_{t,t',p,p'}\right)\right]^2-2\overline{R}^d_{s^*,t}\sum_{(t,t',p,p')\in A_d}\left(p_{s^*}\cdot y^d_{t,t',p,p'}\right)+\overline{R}^{d\ 2}_{s^*,t}\\ &= \sum_{(t,t',p,p')\in A_{sts}(d)}\left(p_{s^*}{}^2\cdot y^d_{t,t',p,p'}\right)-2\overline{R}^d_{s^*,t}\sum_{(t,t',p,p')\in A_d}\left(p_{s^*}\cdot y^d_{t,t',p,p'}\right)+\overline{R}^{d\ 2}_{s^*,t}\\ &= \sum_{(t,t',p,p')\in A_{sts}(d)}\left(p_{s^*}{}^2-2p_{s^*}\cdot\overline{R}^d_{s^*,t}\right)\cdot y^d_{t,t',p,p'}+\overline{R}^{d\ 2}_{s^*,t} \end{aligned} \tag{23}$$

After linearization of the quadratic terms, the subproblem $\boldsymbol{P_y}(d)$ for a single direction d is as follows:

$$min\boldsymbol{P_y}(d) = \sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_{sts}(d)}\left(\widehat{c}^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\times y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right)+C \tag{24}$$

Subject to Constraints (4)-(12) for each direction d. Here, C represents the summation of constant terms, while $\widehat{c}^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}$ represents the general cost for direction d using the STS arc $(t,t',\boldsymbol{p},\boldsymbol{p}')$, as calculated by Equation (25).

$$\widehat{c}^d_{t,t',\boldsymbol{p},\boldsymbol{p}'} = c_w\bullet(t'-t)\bullet\sum_{s\in S(d)}\left(F_t(s)+0.5\sum_{tt=t}^{t'}\alpha_{tt}(s)\right)+c^{\boldsymbol{p}}_d+c_{jd}\bullet\delta^{\boldsymbol{p}}_d+\sum_{s^*\in S^*(d)}\left(\lambda_{d,s^*,t}\bullet p_{s^*}\right)+0.5\rho\sum_{s^*\in S^*(d)}\left(p_{s^*}{}^2-2p_{s^*}\bullet\overline{R}^d_{s^*,t}\right) \tag{25}$$

The subproblem $\boldsymbol{P_y}(d)$ can be regarded as a time-dependent shortest path searching problem, which is easy to solve by our

customized forward DP algorithm (in Appendix A) in a polynomial time.

To provide a better understanding of the time complexity of our customized forward dynamic programming algorithm, let us define the number of time nodes as $|T|$, where each time node $t \in T$ is connected to $|E|$ adjacent time nodes ($|E|$ represents the outgoing arcs from time node t). The value of $|E|$ depends on the minimum and maximum headway, as well as feasible bus formations. Additionally, we define the number of stops on the route as $|S|$. Based on these definitions, we calculate the time complexity as follows:

1. Initializing the label of each time node takes $O(|T|)$ time;
2. Because time only progresses forward in our space–time-state network, we can activate each time node $t \in T$ from the start time t_0 to the end time $|T|$ one by one;
3. Finding and updating the label of each adjacent time node to the time node t takes $O(E)$ time;
4. The loop for the dynamic evolution of passengers takes $O(|S|)$ time;
5. Consequently, from step 3 and step 4 above, the time complexity for updating all the adjacent time nodes to the time node t is $O(|E| \bullet |S|)$;
6. Therefore, the overall time complexity for all time nodes is $O(|T| \bullet |E| \bullet |S|)$;
7. Backtracking the shortest path takes $O(|T|)$ time.

Finally, the overall time complexity of our customized forward dynamic programming algorithm becomes $O(|T|) + O(|T| \bullet |E| \bullet |S|) + O(|T|)$, which is $O(|T| \bullet |E| \bullet |S|)$.

4.1.2. Modular bus unit scheduling subproblem

The sub-model for the modular bus unit scheduling subproblem $\boldsymbol{P_x}$ in our iterative approach is as follows:

$$min\boldsymbol{P_x} = \sum_{d\in D}\sum_{s^*\in S^*(d)}\sum_{t\in T}\lambda_{d,s^*,t}\left[\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_{sts}(d)}\left(p_{s^*}\bullet\overline{y}^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right)-\sum_{a\in\Psi^d_{s^*,t}}\sum_{k\in K}x^k_a\right]+0.5\rho\sum_{d\in D}\sum_{s^*\in S^*(d)}\sum_{t\in T}\left\|\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_{sts}(d)}\left(p_{s^*}\bullet\overline{y}^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right)-\sum_{a\in\Psi^d_{s^*,t}}\sum_{k\in K}x^k_a\right\|_2^2 \tag{26}$$

Subject to Constraints (13) and (14).

This subproblem, which is a quadratic programming (QP) model, can be solved directly by the state-of-the-art commercial solver, Gurobi.

Algorithm 1 shows the procedure of the ADMM-based solution framework to solve our integrated model. In this algorithm, we first assume the initial time–space network for modular bus unit scheduling to be empty, and solve the decomposed subproblems iteratively (see Step 2). Then, the satisfaction of the relaxation Constraint (15) is checked and the upper bound is updated. In Step 4, the Lagrangian multipliers and the penalty parameter are updated at each iteration. Furthermore, motivated by He et al. (2000), we also developed a heuristic adjustment method for the penalty parameter per iteration based on the iterate message in Equation (27). This method helps to improve the convergence speed of the ADMM algorithm in our specific problem setting. As shown in Equation (28), $violation_\rho$ refers to the total number of vehicle resource consistency constraint violations in Constraint (15). Specifically, we set $\rho^0 = 10$, $\beta = 0.25$ and $\sigma = 0.5$.

$$\rho^{n+1} = \begin{cases} \rho^n + \sigma, \|violation_\rho^{n+1}\|_2^2 \geq \beta\|violation_\rho^n\|_2^2 \\ \rho^n - \sigma, \|violation_\rho^{n+1}\|_2^2 = 0 \\ \rho^n, otherwise \end{cases} \tag{27}$$

$$violation_\rho = \sum_{d\in D}\sum_{s^*\in S^*(d)}\sum_{t\in T}\left[\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_{sts}(d)}\left(p_{s^*}\bullet y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right)-\sum_{a\in\Psi^d_{s^*,t}}\sum_{k\in K}x^k_a\right] \tag{28}$$

Input:

Modular bus service STS network for each direction $d \in D$;

Modular bus unit scheduling time-space network for each depot $k \in K$;

The planning horizon $[0, |T|]$ and the time-dependent passenger demand $\alpha_t(s)$ within$[0, |T|]$;

The initial available modular bus units v_k at depot $k \in K$ and the feasible bus formation P;

The maximum number of iterations M and penalty parameters.

Output:

Output the best timetable and bus formation plan $\left\{y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'}\right\}^*$;

Output the best modular bus unit scheduling $\left\{x^k_{i,j,t,t'}\right\}^*$;

Output the best upper bound UB^*.

Step 1: Initialization

Initialize the iteration number $m = 0$;

Initialize the Lagrangian multiplier $\lambda^0_{d,s^*,t'}$ and the penalty parameter ρ^0;

Set the best feasible solution (upper bound) $UB^* = \infty$;

Initialize the vehicle flow variables $\{x^k_a = 0\}$ for all the time-space arcs in the modular bus unit scheduling network for each depot $k \in K$, and compute $\bar{R}^d_{s^*,t} = \sum_{a\in\Psi^d_{s^*,t}} \sum_{k\in K} \bar{x}^k_a$ as input for subproblem $P(y)$.

Step 2: Solving the decomposed subproblems iteratively

Step 2.1 Solving the timetable and bus formation plan subproblem $P(y)$

For *each direction* $d \in D$ **do**

Call the forward DP algorithm, as described in Appendix A, to find the shortest path from vertex $(t_0, \boldsymbol{p_0})$ to $(|T|,*)$;

Let $y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'} = 1$ for arc $(t, t', \boldsymbol{p}, \boldsymbol{p}')$ along the shortest path;

Compute $\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_{sts}(d)} \left(p_{s^*} \cdot \bar{y}^d_{t,t',\boldsymbol{p},\boldsymbol{p}'} \right)$ $\forall d \in D, \forall s^* \in S^*(d), \forall t \in T$, as input for subproblem $P(x)$;

End

Step 2.2 Solving the modular bus unit scheduling subproblem $P(x)$

Call the Gurobi solver to solve the modular bus unit scheduling subproblem $P(x)$;

Compute $\bar{R}^d_{s^*,t} = \sum_{a\in\Psi^d_{s^*,t}} \sum_{k\in K} \bar{x}^k_a$ as input for subproblem $P(y)$ in the next iteration.

Step 3: Update the best upper bound value

if *the ADMM solution satisfies all the relaxation constraints in Constraint (15)* **then**

Calculate the upper bound value UB^m by Equation (19) and update the best upper bound value by $UB^* = \{UB^*, UB^m\}$

End

Step 4: Update the Lagrangian multipliers and penalty parameters

Update the value of the Lagrange multipliers $\lambda^{m+1}_{d,s^*,t}$ in the $(m+1)th$ iteration using the following equation:

$$\lambda^{m+1}_{d,s^*,t} = \lambda^m_{d,s^*,t} + \rho^m \left(\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}')\in A_{sts}(d)} \left(p_{s^*} \cdot y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'} \right) - \sum_{a\in\Psi^d_{s^*,t}} \sum_{k\in K} x^k_a \right)$$

Update penalty parameter ρ^{m+1} in the $(m+1)th$ by Equation (27)

Step 5: Terminal conditions

if $m < M$, let $m = m + 1$, and go back to step 2; otherwise, output the best upper bound solution and terminate the algorithm.

4.2. LR-based solution framework

The solution derived from the ADMM-based solution framework constitutes a feasible upper bound on the optimal solution. To assess the quality of this feasible solution, we construct a pure LR model, which yields a corresponding lower bound solution. To achieve this, we eliminate the quadratic term from Equation (21), resulting in the following LR relaxation model:

$$minZ_L = \sum_{d\in D}\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)} y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\left[c_w \bullet (t^{'}-t)\bullet \sum_{s\in S(d)}\left(F_t(s)+0.5\sum_{tt=t}^{t^{'}}\alpha_{tt}(s)\right)+c^{\boldsymbol{p}}_{d}+c_{jd}\bullet\delta^{\boldsymbol{p}}_{d}\right]+\sum_{d\in D}\sum_{s^{*}\in S^{*}(d)}\sum_{t\in T}\underline{\lambda_{d,s^{*},t}}\left[\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)}\left(p_{s^{*}}\bullet y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\right)-\sum_{a\in\Psi^{d}_{s^{*},t}}\sum_{k\in K}x^{k}_{a}\right] \tag{29}$$

Subject to Constraints (4)-(14).

Similar to the decomposition process of the ADMM-based model, the LR-based model can also be decomposed into two subproblems based on two independent sets of variables.

Regarding the subproblem $\underline{\boldsymbol{P}_{\boldsymbol{y}}}$, it addresses the timetable and bus formation plan, and is presented below:

$$min\underline{\boldsymbol{P}_{\boldsymbol{y}}} = \sum_{d\in D}\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)} y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\left[c_w \bullet (t^{'}-t)\bullet \sum_{s\in S(d)}\left(F_t(s)+0.5\sum_{tt=t}^{t^{'}}\alpha_{tt}(s)\right)+c^{\boldsymbol{p}}_{d}+c_{jd}\bullet\delta^{\boldsymbol{p}}_{d}\right]+\sum_{d\in D}\sum_{s^{*}\in S^{*}(d)}\sum_{t\in T}\underline{\lambda_{d,s^{*},t}}\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)}\left(p_{s^{*}}\bullet y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\right) \tag{30}$$

Subject to Constraints (4)-(12).

Like the second-level decomposition of the $\boldsymbol{P}_{\boldsymbol{y}}$ in ADMM-based model, the subproblem $\underline{\boldsymbol{P}_{\boldsymbol{y}}}$ in the LR-based model can also be further decomposed into subproblems with respect to individual directions, as shown follows:

$$min\underline{\boldsymbol{P}_{\boldsymbol{y}}(d)} = \sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)} y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\left[c_w \bullet (t^{'}-t)\bullet \sum_{s\in S(d)}\left(F_t(s)+0.5\sum_{tt=t}^{t^{'}}\alpha_{tt}(s)\right)+c^{\boldsymbol{p}}_{d}+c_{jd}\bullet\delta^{\boldsymbol{p}}_{d}\right]+\sum_{s^{*}\in S^{*}(d)}\sum_{t\in T}\underline{\lambda_{d,s^{*},t}}\sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)}\left(p_{s^{*}}\bullet y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\right) \tag{31}$$

Subject to Constraints (4)-(12) for direction d.

The objective function in Equation (31) can be reformulated as Equation (32) as follows:

$$min\underline{\boldsymbol{P}_{\boldsymbol{y}}(d)} = \sum_{(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})\in A_{sts}(d)}\left(\overline{c}^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\bullet y^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}\right)+C \tag{32}$$

Subject to Constraints (4)-(12) for direction d. Here, C represents the summation of constant terms, while $\overline{c}^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}}$ represents the general cost for direction d using the STS arc $(t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'})$, as calculated by Equation (33).

$$\overline{c}^{d}_{t,t^{'},\boldsymbol{p},\boldsymbol{p}^{'}} = c_w \bullet (t^{'}-t)\bullet \sum_{s\in S(d)}\left(F_t(s)+0.5\sum_{tt=t}^{t^{'}}\alpha_{tt}(s)\right)+c^{\boldsymbol{p}}_{d}+c_{jd}\bullet\delta^{\boldsymbol{p}}_{d}+\sum_{s^{*}\in S^{*}(d)}\left(\underline{\lambda_{d,s^{*},t}}\bullet p_{s^{*}}\right) \tag{33}$$

The subproblem $\underline{\boldsymbol{P}_{\boldsymbol{y}}(d)}$ can be viewed as a time-dependent shortest path search problem that can be efficiently solved using DP in polynomial time. The DP algorithm used to solve this subproblem $\underline{\boldsymbol{P}_{\boldsymbol{y}}(d)}$ are similar to $\boldsymbol{P}_{\boldsymbol{y}}(d)$ in the ADMM-based solution procedure. The only difference is the LR cost for each STS arc. We use the traditional sub-gradient method to update multipliers $\underline{\lambda_{d,s^{*},t}}$ in LR iterations.

Regarding the subproblem $\underline{\boldsymbol{P}_{\boldsymbol{x}}}$, it addresses the modular bus unit scheduling, and is presented below:

$$min\underline{\boldsymbol{P}_{\boldsymbol{x}}} = -\sum_{d\in D}\sum_{s^{*}\in S^{*}(d)}\sum_{t\in T}\underline{\lambda_{d,s^{*},t}}\sum_{a\in\Psi^{d}_{s^{*},t}}\sum_{k\in K}x^{k}_{a} \tag{34}$$

Subject to Constraints (13) and (14).

Subproblem $\underline{\boldsymbol{P}_{\boldsymbol{x}}}$, which is an integer linear programming model, can be solved directly by the state-of-the-art commercial solver, Gurobi. Our LR problem (Z_L) is solved by a sub-gradient method, and the detailed procedure is given in Algorithm 3 (Appendix B).

5. Numerical experiments

This section presents numerical experiments that investigate the validity of the integrated model and the computation performance of our proposed ADMM-based solution framework. To evaluate the performance of our model and solution framework, we generate various scale instances based on a hypothetical bidirectional bus line. Furthermore, we compare three operational strategies using real-world data from bus line 300 in Beijing, China, to assess the effectiveness of the proposed AMV-based operational strategy and flexible vehicle scheduling. This real-world numerical experiment also evaluates the applicability of our proposed ADMM-based solution framework in solving real-world cases. All programs are implemented in Python and executed on an Intel Core i7-11700 CPU operating at 2.50 GHz with 32 GB RAM. We use Gurobi version 9.5.2 to conduct the experiments.

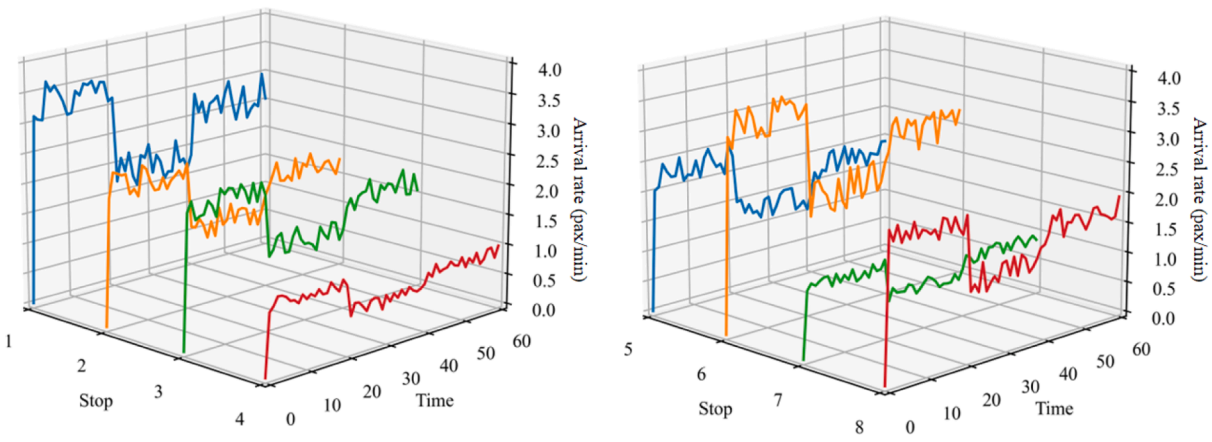
5.1. Illustrative experiments

5.1.1. Setup

The purpose of several illustrative experiments is to investigate the validity of our proposed model and the computation performance of our developed ADMM-based solution framework. The length of the minimum time interval δ is set to 1 min in both the STS network for modular bus service and the time–space network for modular bus unit scheduling. We consider a hypothetical bidirectional loop line, as illustrated in Fig. 5, where the distance between adjacent stops is equal, and the running time is 2 min. The travel time for modular bus units between the depots and intermediate stops via outgoing/incoming depot arcs is 1 min. The time-dependent passenger arrival rate for each stop on two directions is shown in Fig. 11. We set the default value of parameters in the AMV-based PT system as follows: $P = [1, 2, 3]$, $Cap = 6$ pax/modular bus unit. The unit-waiting cost per passenger per minute c_w is set as 0.8$/min, according to Chen et al. (2019). For the operational costs of a modular bus, the fixed operational costs C^F is set as 1.912$, while the marginal operational costs C^V is set as 0.59$ per seat, according to Chen and Li (2021). Therefore, the total operational costs of a modular bus in formation p is considered as: $c^p = 1.912 + 0.59(p \bullet Cap)^{\alpha}$, in which the parameter α was set to 1. The penalty cost for a detachment or joining operation c_{jd} is set as 1.5$. Additionally, the minimum headway h_{min} and maximum headway h_{max} are set as 2 min and 8 min, respectively. To ensure both solution efficiency and accuracy, we set the maximum number of iterations M for the ADMM-based solution framework to 100 in all our numerical experiments. We generate cases with different planning horizons and number of available modular bus units based on the hypothetical bus line. To evaluate the computation performance of the proposed ADMM-based solution framework, we compare its performance with a state-of-the-art commercial solver, Gurobi. The superiority of the integrated optimization method over the sequential optimization method is also demonstrated by experimental examples.

5.1.2. Optimization results

Table 4 displays the detailed computational results for the various test cases conducted under different scenarios. Each case is named after $|T|/v$, where $|T|$ denotes the planning horizon (in minutes), and v represents the number of available modular bus units at each terminal depot. In Table 4, most instances demonstrate the effectiveness of the proposed ADMM-based solution framework by efficiently producing very good quality solutions with low gap. The optimality gap is determined by calculating the difference between the lower bound value obtained from the pure LR algorithm and the upper bound value, expressed as a percentage $(UB - LB)/UB \times 100\%$. However, it is worth noting that one instance deviates from this trend, exhibiting an outlier with an optimality gap of 8.67%. Nevertheless, this value remains below the 9% threshold, signifying a relatively satisfactory performance. This particular case corresponds to scenario $v = 3$, characterized by a significantly limited number of available modular bus units. The scarcity of vehicle resources in this scenario severely restricts the solution space, thereby constraining the algorithm's search capability and resulting in diminished performance. With an increase in the planning horizon, the computational time for Gurobi (with a maximum computation time of 3600 s) is substantially higher compared to the ADMM-based solution framework. Moreover, as shown in Table 5, Gurobi is unable to produce a feasible solution when the planning horizon is larger than 10 min, as it cannot even read the model with an extremely large number of variables and constraints. In contrast, the effectiveness of our proposed ADMM-based solution framework is apparent. When compared to the Gurobi solver, it not only significantly reduces computation time but also obtains better feasible solutions. As shown in Fig. 12 (a), the computation time of our ADMM-based solution framework increases linearly with the planning horizon, indicating its potential applicability to large-scale problems. Fig. 12 (b) illustrates that the computation time of the ADMM-based solution framework is nearly independent of the number of available modular bus units, highlighting the advantage of our flow-based vehicle scheduling model in a time–space network. However, a significant limitation of our solution framework is that the optimality gap is relatively larger when the number of available modular bus units is scarce. This is likely because obtaining a smaller



(a) Passenger arrival rate on clockwise route

(b) Passenger arrival rate on counter-clockwise route

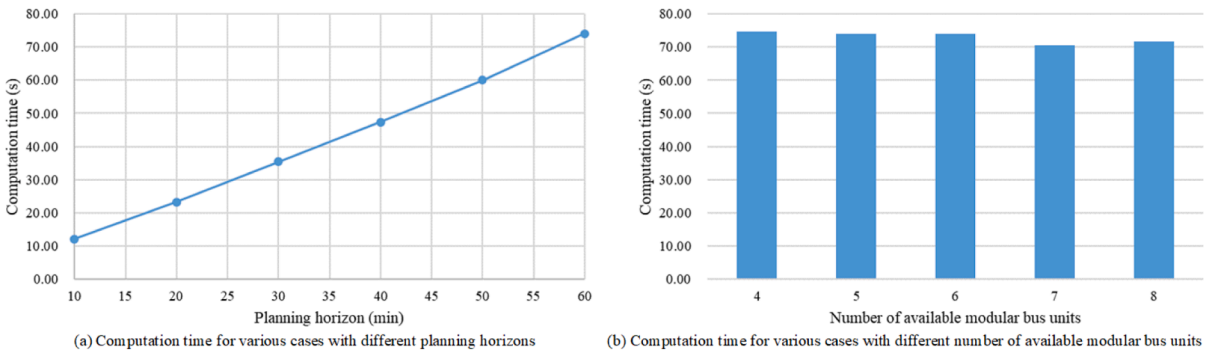
Fig. 11. Temporal and spatial patterns of passenger demand.

Table 4
Detailed computational results for the various test cases.

Scenario	Computation time (s)		Objective value ($)		Optimality gap (%)	Modular bus unit usage	Number of deadhead trip
	Integrated optimization	Sequential optimization	Integrated optimization	Sequential optimization			
10/5	12.15	0.01	358.15	358.15	0.33	[5, 5]	0
20/5	23.29	–	798.87	–	0.60	[5, 5]	2
30/5	35.45	–	1078.24	–	0.27	[5, 5]	2
40/5	47.36	–	1349.38	–	0.20	[5, 5]	2
50/5	60.03	–	1696.00	–	0.91	[5, 5]	5
60/4	74.60	–	2319.86	–	8.67	[4, 4]	6
60/5	74.00	–	2002.37	–	0.23	[5, 5]	4
60/6	73.93	0.57	1987.98	1987.98	0.08	[6, 6]	2
60/7	70.47	0.57	1987.98	1987.98	0.08	[7, 5]	1
60/8	71.62	0.58	1987.98	1987.98	0.54	[7, 5]	1

Table 5
Comparison of ADMM-based solution framework and Gurobi.

Scenario	Number of variables	Number of constraints	Solution approach	Computation time (s)	Best upper bound value ($)	Best lower bound value ($)	Optimality gap (%)
8/5	199,858	1,585,806	Gurobi	3600	276.44	267.96	3.07
			ADMM	8.66	276.44	275.40	0.38
10/5	370,886	2,952,634	Gurobi	3600	378.34	330.74	12.58
			ADMM	12.15	358.15	356.97	0.33
12/5	554,874	4,423,142	Gurobi	3600	–	402.02	–
			ADMM	12.45	449.98	443.31	1.48



(a) Computation time for various cases with different planning horizons
(b) Computation time for various cases with different number of available modular bus units

Fig. 12. Variation of computation time with different key model parameters.

Table 6
The impact of available modular bus units.

Planning horizons (min)	Number of available modular bus units	Total passenger waiting cost ($)	Operational costs ($)	Detachment and joining cost ($)	Total system costs ($)
60	3	–	–	–	–
	4	1279.40	939.58	18	2236.98
	5	958.07	1020.30	24	2002.37
	6	911.80	1049.19	27	1987.98
	7	911.80	1049.19	27	1987.98
	8	911.80	1049.19	27	1987.98

upper bound or a larger lower bound or both is more challenging. Nevertheless, cases of extreme vehicle resource scarcity are rare. As shown in Table 6, if the number of available modular bus units at each terminal depot is one less, the model will not obtain a feasible solution, implying that the waiting passengers at stops cannot be served at the end time of the planning horizon. Although this scenario is seldom encountered in reality, it is common for vehicle scheduling to fail to cover the optimal timetable, and we compare the integrated model with sequential optimization in Section 5.1.3.

5.1.3. Comparison with sequential optimization

This section presents the results of our experiments, which compare the integrated optimization of timetable and bus formation with vehicle scheduling against a sequential optimization approach that follows the hierarchical planning process of public transport. In the comparison, the sequential optimization approach firstly employs our tailored forward dynamic programming to address the timetable and bus formation subproblem. Then, the calculated timetable and bus formation plan serve as inputs for the subsequent vehicle scheduling subproblem. The results shown in Table 4 indicate that the sequential optimization approach, with planning horizons ranging from 20 to 60 min and 5 modular bus units at each terminal depot, fails to produce a feasible solution due to an inadequate number of modular bus units. This finding suggests that simply balancing the costs of passengers and operators during the timetabling stage is insufficient when passenger demand exceeds transportation supply, especially during peak periods. As demonstrated in Table 6, a shortage of modular bus units leads to a significant increase in passengers' waiting costs. From an operator's perspective, optimizing the timetable and bus formation while cooperating with modular bus unit scheduling is therefore crucial in mitigating these costs.

5.1.4. The impact of available modular bus units

We conducted a numerical experiment to examine the impact of available modular bus units on modular bus service and modular bus unit scheduling. Using the illustrative experiment from Section 5.1.1 as a basis, we varied the number of available modular bus units at each terminal depot while maintaining a fixed planning horizon of 60 min, and observed the corresponding impact on optimization results. Table 4 and Table 6 illustrate that when modular bus units are scarce, the total system costs and the number of deadhead trips increase, and the timetable and bus formation plan cannot achieve optimal results. This suggests that sequential optimization is ineffective under these conditions. Conversely, when sufficient modular bus units are available to meet the optimal timetable and bus formation plan, an increase in available modular bus units cannot reduce total costs but can reduce the number of deadhead trips by optimizing the usage of modular bus units at terminal depots.

In particular, when modular bus units are scarce, operators cannot schedule all trips optimally due to vehicle resource constraints. Consequently, some trips may have to be scheduled at suboptimal times or canceled, leading to a substantial increase in the total passenger waiting time cost. Additionally, scarce modular bus unit resources increase the chance of deadhead trips, resulting in energy waste due to hauling empty units. However, when there are sufficient modular bus units available, increasing their number can reduce the number of deadhead trips by scheduling the same modular bus units to depart from different terminal depots. As a result, the PT system can reduce the number of empty units, thereby decreasing energy waste.

5.1.5. The impact of detachment and joining penalty cost

This section presents the results of a numerical experiment that evaluated the impact of the penalty cost associated with a detachment or joining operation on scenario 60/6 in Section 5.1.1. The findings demonstrate the impact of different values of penalty cost on the passengers' and the operator's costs, and the total system costs. As shown in Table 7, results reveal that as the penalty cost increases from 1.5 to 12, the total system costs also increase, while the total passenger waiting time cost decreases, and the operational costs increase. Moreover, the total number of en-route detachment and joining operations decreases. If the penalty cost exceeds 12, the number of detachment and joining operations becomes zero, resulting in a degeneration from fully flexible capacity to semi-flexible capacity operations, and the modular bus can only adjust its formation at terminal depots. Besides, further increasing the penalty cost does not impact the total system costs. We note that once AMV technology becomes sufficiently developed and tested to ensure safe operations, the penalty cost can be eliminated. This would enable the best trade-off between passengers' and operational costs to achieve optimal total system costs. Therefore, the usage of such a penalty depends on the maturity of the technology, where advanced vehicles in the future will be able to do these operations risk-free.

5.2. Line 300 experiments in Beijing public transport system

5.2.1. Network description and parameter setting

In this section, we present a real-world numerical experiment that assesses the potential of introducing our proposed AMV-based PT system in Line 300 of the Beijing Public Transport System. Line 300 is a vital bus line in Beijing, encircling the city's core areas and providing dedicated bus lanes during peak periods to ensure punctual service. The line comprises 26 stops and two terminal depots, as shown in Fig. 13. Besides, two intermediate depots are strategically located along this loop line to temporarily store vehicles serving Line 300. For practical application, the selected assemblage stops for this experiment are located near existing depots that are already

Table 7
The impact of penalty cost for detachment and joining cost.

Penalty cost ($)	Total passenger waiting cost ($)	Operational costs ($)	Detachment and joining cost ($)	Total system costs ($)	Number of en-route detachment and joining operations
0	902.95	1041.54	0.00	1944.49	27
1.5	911.80	1049.19	27.00	1987.98	18
6	892.89	1111.64	24.00	2028.53	4
10.5	887.62	1125.80	21.00	2034.42	2
12	887.62	1147.04	0.00	2034.66	0
13.5	887.62	1147.04	0.00	2034.66	0

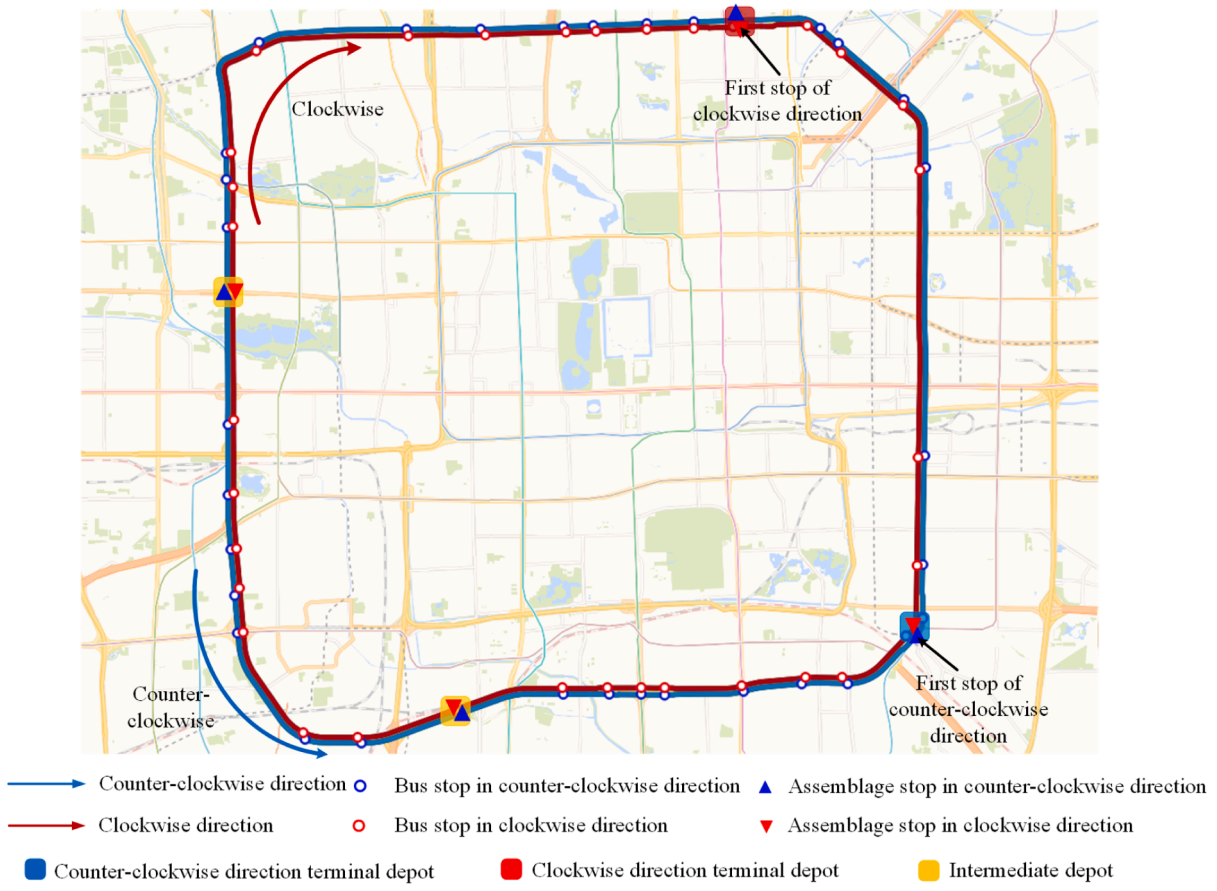


Fig. 13. Line 300 of the Beijing Public Transport System.

in operation. Moreover, short-turning strategies are commonly used on bus lines, including Line 300. Therefore, scheduling modular bus units in both directions simultaneously and allowing them to detach from one modular bus and join another modular bus in either direction is feasible.

The study period covers the period from 6:00–11:00 on a typical weekday, which includes both peak and off-peak periods, as shown in Fig. 1(a), giving obvious temporal fluctuations in passenger demand. We set the length of the minimum time interval δ to 1 min in both modular bus service STS network and modular bus unit scheduling time–space network. We adopt the following default parameter values: $Cap = 30$ pax/modular bus unit, $v = 54$, $h_{min} = 2min$, $h_{max} = 10min$, $c_{jd} = 10\$$. The remaining parameter values remain unchanged from those specified in Section 5.1.1.

5.2.2. *Operational strategy comparison*

We firstly evaluate the efficacy of our proposed AMV-based PT system and examine the applicability of the ADMM-based solution framework in solving real-world cases. The values of the best upper bound and best lower bound evolved with the number of iterations, as shown in Fig. 14. It is observed that the solution can be improved multiple times during the algorithm's runtime and eventually achieves an acceptable optimality gap.

Fig. 15 depicts the derived timetable and bus formation plan for each direction. It reveals that dispatching headways during peak hours, specifically between 7:00 and 9:00, are significantly shorter than those observed between 9:00 and 11:00. This responsiveness to time-dependent demand results in a higher frequency of modular bus dispatches during peak periods. Moreover, larger vehicle formations for modular buses are primarily reserved for peak periods. However, due to spatial fluctuations in passenger demand along the route, modular buses do not have to maintain such formations for the entire route. Overall, the findings suggest that larger headways and smaller vehicle formations of modular buses are suitable for low demand levels, whereas shorter headways and larger vehicle formations of modular buses are preferred when demands are higher. These results align with the inherent characteristics of modular buses, allowing them to respond to spatiotemporal fluctuations in passenger demand.

Next, we compared our proposed AMV-based system with two benchmark operational strategies which have been discussed in Section 1.1. The first one is the fixed-capacity operational strategy where the fixed-capacity bus comprises multiple modular bus units combined and is no longer separated. The second one is the semi-flexible operational strategy where modular buses are allowed to

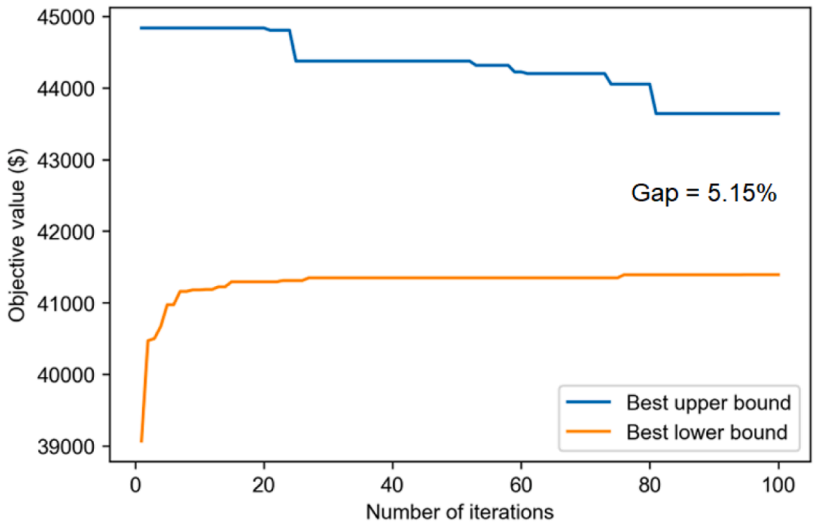


Fig. 14. Evolution of best upper bound and best lower bound.

change their formations to $p = [1, 2, 3]$ in the terminal depots. The overall system costs, vehicle operational costs, passenger waiting time cost, and detachment and joining cost are all used to assess the operational strategy performance. The comparision of the three different operation scenarios is summarized in Table 8 and the details of the three scenarios are described as follows.

Scenario 1: This scenario encompasses three cases in which the fixed-capacity bus can accommodate $p = 1,2,3$, corresponding to capacities equivalent to 30, 60, and 90 passengers per vehicle, respectively. In this scenario, operators can only optimize the dispatching time and scheduling of available buses synchronously to meet the demand, resembling the current operation of public transport. For each case, there are 54, 27, and 18 buses available for utilization at each terminal depot, respectively.

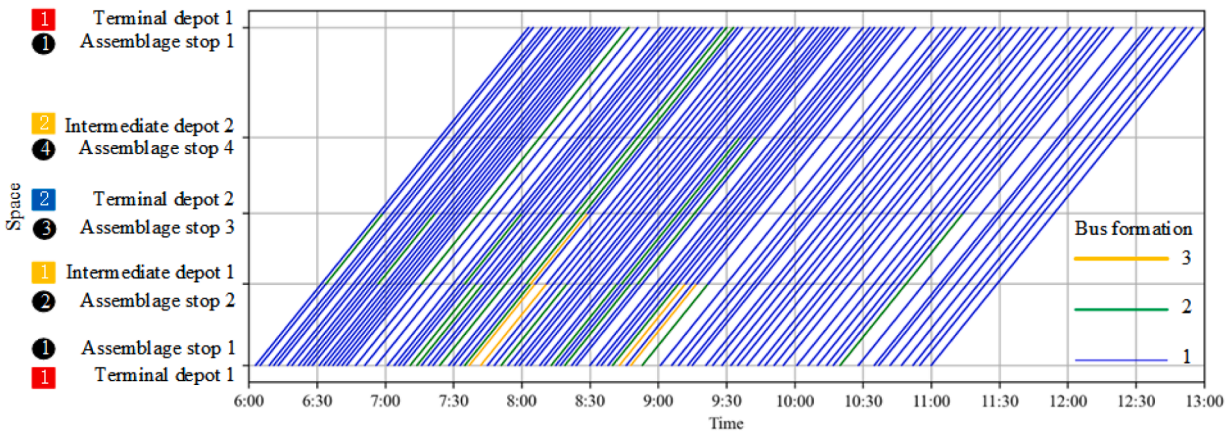
Scenario 2: In this scenario, we have the opportunity to enhance not only the efficiency of dispatching time and scheduling of modular bus units but also the optimization of bus formation. Under this scenario, modular buses are allowed to change their formations p in the terminal depots, where $p = [1, 2, 3]$ corresponding to the capacity $Cap = [30, 60, 90]$. There are 54 modular bus units available for usage at each terminal depot at the beginning.

Scenario 3: This is the operational strategy that our study focusses on in Section 5.2.1.

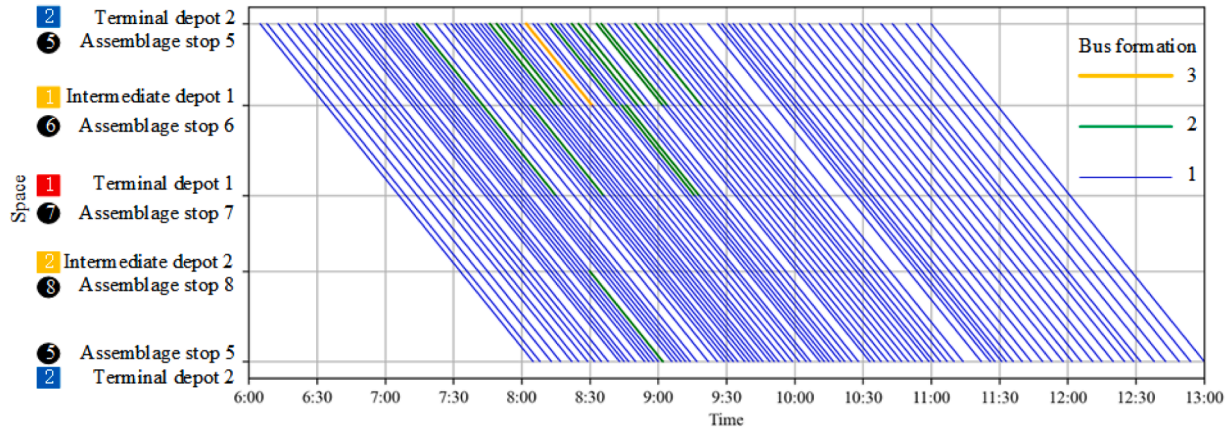
Table 9 shows the performance comparison results of different scenarios. Taking Scenario 1 ($p = 3$) as the benchmark, we observed that the introduction of AMV technology into the PT system could reduce the overall system costs. In particular, both operational strategies in Scenario 2 and Scenario 3 reduce passenger waiting time costs by more than 60%. However, the higher vehicle operational costs in Scenarios 2 and 3 than in Scenario 1 was an unexpected finding. This is because when trading-off operational costs and passenger waiting time cost, Scenarios 2 and 3 require less vehicle operational costs to reduce passenger waiting time cost by dispatching buses with specific formations to match the time-dependent passenger demand. In contrast, the operational strategy in Scenario 1 requires more cost to dispatch a fixed capacity bus, so it prefers to dispatch buses when a certain amount of passengers have accumulated at the stops. Therefore, more frequent bus dispatching in Scenarios 2 and 3 results in higher vehicle operational costs than in Scenario 1.

As depicted in Fig. 14, the optimized timetable and bus formation plan predominantly utilize bus formations with $p = 1$. By employing smaller buses, the system can achieve a significantly reduced headway, which greatly reduces passenger waiting time. Therefore, comparing the fixed-capacity operational strategy using large buses in Scenario 1 ($p = 3$) with the modular bus operational strategies in Scenarios 2 and 3 is inherently unfair. To demonstrate that the decrease in total system costs is attributable to the modular bus strategy, rather than solely the reduction in headway achieved with small bus formations, we also introduced Scenario 1 with small buses in both $p = 2$ and $p = 1$. The results in Table 9 indicate that the smaller bus formation in Scenario 1 can decrease the total system costs, but Scenario 3 outperforms Scenario 1 ($p = 1$) with even lower total system costs. This indicates that Scenario 1 ($p = 1$) can only dispatch small-capacity buses at shorter headways to decrease total passenger waiting time. However, even though autonomous buses do not incur labor costs, there are fixed costs associated with dispatching a bus. Consequently, despite the slightly smaller total passenger waiting time cost in Scenario 1 ($p = 1$) compared to Scenario 3, the operational costs of Scenario 1 ($p = 1$) are 11.16% higher than those of Scenario 3. Therefore, the modular bus operational strategy offers a superior trade-off between passengers' and operators' costs, resulting in decreased total system costs when compared to the fixed-capacity operational strategy.

In Scenario 3, allowing modular buses to change their vehicle formations along the route enable the system to better match spatial fluctuation of passenger demand compared to Scenario 2. Therefore, the operational strategy in Scenario 3 can reduce vehicle operational costs by using short formations of buses on low passenger load sections on the route. However, the total system costs are not significantly lower than those in Scenario 2 because of the additional detachment and joining penalty cost. This penalty cost is directly related to the level of AMV technology development. Therefore, operators need to carefully consider the penalty cost of en-



(a) Timetable and bus formation plan of clockwise direction under fully flexible capacity operations



(b) Timetable and bus formation plan of counter-clockwise direction under fully flexible capacity operations

Fig. 15. Optimal timetable and bus formation in two directions of Line 300.

Table 8
Comparison of three different operation scenarios.

Scenario	Whether synchronously optimizing timetable and vehicle scheduling?	Whether synchronously optimizing bus formation?	Whether to allow a modular bus to detach and join en route?
1	√	×	×
2	√	√	×
3	√	√	√

Table 9
Computation results of the three different operation scenarios.

Scenario	Total passenger waiting cost ($)	Operational costs ($)	Detachment and joining cost ($)	Total system costs ($)
1 ($p = 3$)	76827.52	6981.87	0	83809.39
1 ($p = 2$)	43240.17	19612.00	0	62852.17
1 ($p = 1$)	27031.22	17650.80	0	44682.02
2	27214.03	16774.54	0	43988.57
3	27331.35	15878.62	430	43639.97

route detachment and joining operations at different stages of AMV development. This finding is consistent with the numerical experiment in Section 5.1.5.

5.2.3. Vehicle scheduling strategy comparison

The previous section demonstrates that incorporating AMV technology in the PT system can lower overall system costs and improve the balance between operational costs and passenger waiting time cost. However, allowing modular buses to detach and join at intermediate stops in one direction leads to an accumulation of modular bus units at specific intermediate depots, which will result in deadhead trips when balancing modular bus units to other stops. To illustrate the efficacy of the simultaneous vehicle scheduling strategy in both directions, we compared the results with vehicle scheduling independently in one direction, while keeping all parameter values the same as those in Scenario 3.

Fig. 16 presents the scheduling of individual modular bus units under different vehicle scheduling strategies. Red blocks of varying shades represent modular bus units carrying out trips on different segments of the clockwise direction, while blue blocks of varying shades indicate modular bus units carrying out trips on different segments of the counter-clockwise direction. Green blocks denote modular bus units performing deadhead trips. As shown in Fig. 16, a modular bus unit is occupied when it is dispatched for a scheduled trip, and idle otherwise. The total time that modular bus units are occupied decreased by 0.35% under the simultaneous vehicle scheduling strategy, as compared to vehicle scheduling independently in one direction. However, Table 10 indicates a 4.13% reduction in overall system costs for the simultaneous vehicle scheduling strategy compared to separate vehicle scheduling in one direction. Additionally, both strategies utilize all the available modular bus units effectively, suggesting that the flexible vehicle scheduling strategy leads to lower total system costs while maintaining a similar vehicle utilization rate. This is primarily due to the 55% reduction in the number of deadhead trips.

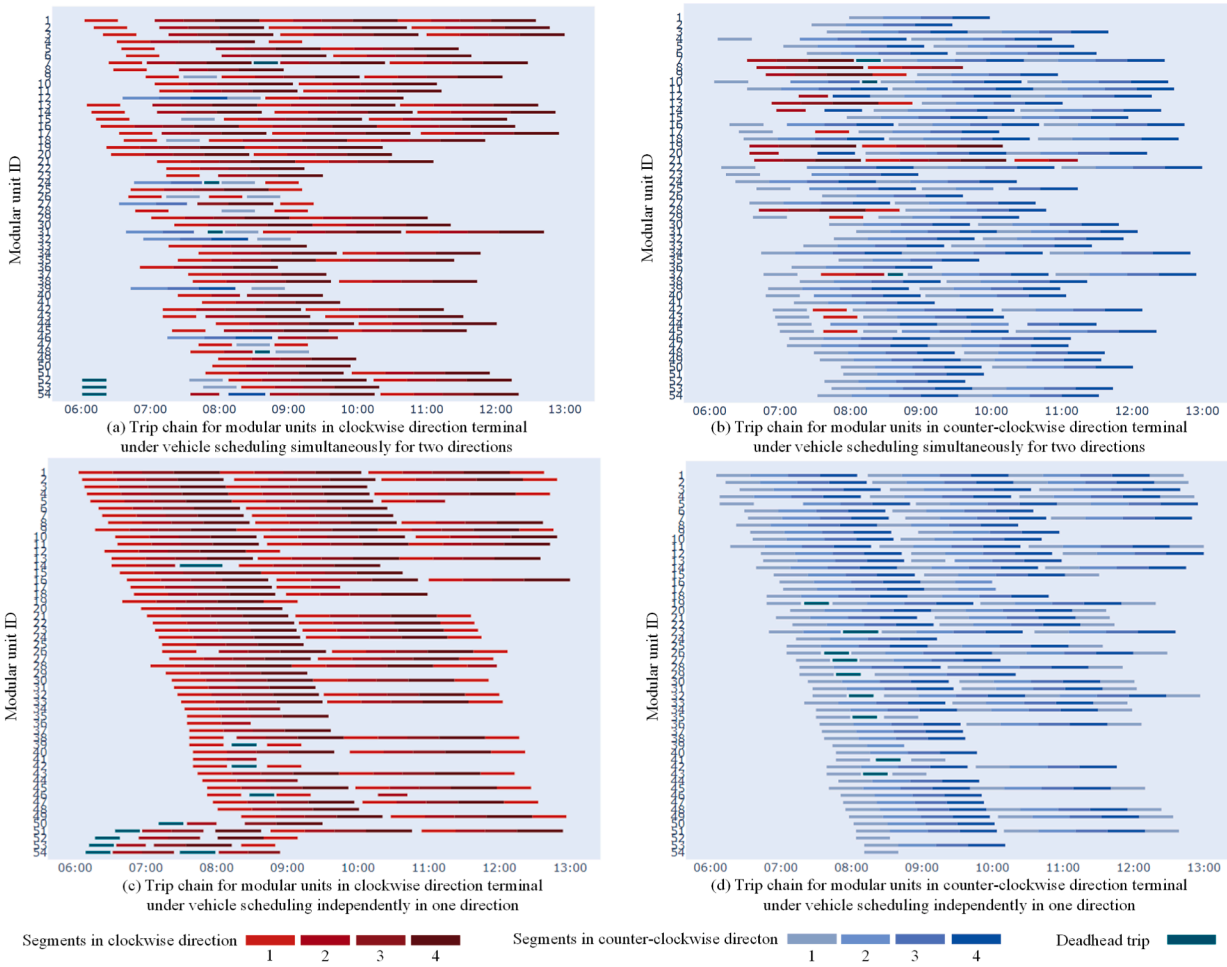


Fig. 16. Modular bus units scheduling under different vehicle scheduling strategies.

Table 10
Computation results of the two different vehicle scheduling strategies.

Vehicle scheduling strategy	Total system costs ($)	Modular bus unit usage	Occupied time (min)	Number of deadhead trips
Simultaneously	43639.97	[54, 54]	24,545	10
Separately	45518.93	[54, 54]	24,631	22

6. Conclusions and future research

In this paper, we presented a joint optimization model for the timetable, bus formation, and vehicle scheduling of a bidirectional bus line that incorporates en-route detachment and joining of autonomous modular bus units. Our model included a penalty cost for each detachment or joining operation, and we formulated the timetable and bus formation problem on a novel STS network, introducing an endogenous rule to reduce redundant arcs. Additionally, we formulated the modular bus unit scheduling problem on a time–space network and proposed a flexible scheduling strategy that allowed for modular bus units to detach from and join other modular buses in either direction. We coupled the above two subproblems with a consistency constraint and formulated an integrated model to solve the problem. Due to the complexity of our model, we have introduced ADMM to decompose it into two easy-to-solve subproblems which can be solved by our customized forward DP algorithm and commercial solver. Our proposed model and solution framework were validated using a set of illustrative examples and real-world instances of Line 300 in Beijing Public Transport. Our proposed ADMM-based solution framework significantly reduced computation time while achieving a comparable solution performance to the state-of-the-art solver, Gurobi. We observed that the AMV-based public transportation system can better trade-off vehicle operational costs and passenger waiting time cost. Moreover, we found that simultaneous vehicle scheduling in both directions can save total system costs and significantly reduce the number of deadhead trips.

We identified several directions for further research. First, the acquisition costs were not considered in our research due to the tactical nature of our problem. In practice, the acquisition costs play a crucial role in conducting comprehensive technical and economic evaluations of AMVs from a strategical perspective, which will be a key focus for future research. Second, our model can easily incorporate electric energy by adding an energy dimension to the vehicle scheduling time–space network to address future transportation electrification. Third, it would be interesting to consider uncertain passenger demand or the performance improvements brought by AMV technology at the public transportation network level (Liu et al., 2021b). Finally, future studies may consider the detachment and joining operations of AMVs at any stops along the route under the scenario of sufficiently developed AMV technology, leading to more efficient operational strategies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This paper is supported by National Natural Science Foundation of China (52072017, 52272420) and Beijing Natural Science Foundation (8212010).

Appendix A

Input:
Modular bus service STS network $A_{sts}(d)$ of direction d;
The planning horizon $[0, |T|]$ and the time-dependent passenger demand OD within$[0, |T|]$;
The feasible bus formation P.

Step 1: Initialization
For time node $t = 1$ to $|T|$ **do**
  $label(t) = +\infty$;
  Time predecessor of the STS node $(t,*) = -1$;
  State predecessor of the STS node $(t,*) = -1$;
End
$label(t_0) = 0$;
Step 2: Forward finding the best node on the shortest path
For time node $t = t_0$ to $|T|$ **do**
  Derive feasible downstream time node t' for t based on Rule 1;
  For state vector $\boldsymbol{p} \in P_d$ **do**
    For stop $s \in S(d)$ **do** // *dynamic evolution of passengers*
      Calculate the number of waiting passengers at stop s at time t $(W_t(s))$;
      Calculate the proportion of passengers at stop s at time t with destination of stop s' $(\beta_t(s,s'))$;
      Calculate the number of alighting passengers from a modular bus at stop s at time t $(A_t(s))$;
      Calculate the number of boarding passengers to a modular bus at stop s at time t $(B_t(s))$;
      Calculate the number of passengers who fail to board a modular bus at stop s at time t $(F_t(s))$;
      Calculate the number of in-vehicle passengers of a modular bus departing from stop s at time t $(I_t(s))$;
    End
    Update the generalized cost $\hat{c}^d_{t,t',\boldsymbol{p},*}$ of the STS arc $(t, t', \boldsymbol{p}, *)$ according to Equation (26);
    If $label(t) + \hat{c}^d_{t,t',\boldsymbol{p},*} \le label(t')$ **then**
      $label(t') = label(t) + \hat{c}^d_{t,t',\boldsymbol{p},*}$;
      Time predecessor of the STS node $(t',*) = t$;
      State predecessor of the STS node $(t',*) = \boldsymbol{p}$;
    End
  End
End
Step 3: Backtrack the shortest path
Find the last STS node $(|T|,*)$ on the shortest path obtained by DP;
Backtrack from the last STS node to the first STS node $(t_0, \boldsymbol{p_0})$ based on the precedence relationship;
Reverse the backward path and output the shortest path and the corresponding cost.

Appendix B

Input:

Modular bus service STS network for each direction $d \in D$;

Modular bus unit scheduling time-space network for each depot $k \in K$;

The planning horizon $[t_0, |T|]$ and the time-dependent passenger demand OD within$[t_0, |T|]$;

The initial available modular bus units v_k at depot $k \in K$ and the feasible bus formation P;

The maximum number of iterations M and penalty parameters.

Output:

Output the best lower bound LB^*.

Step 1: Initialization

Initialize the iteration number $m = 0$;

Initialize the Lagrangian multiplier $\underline{\lambda^0_{d,s^*,t}}$;

Set the LR solution (lower bound) $LB^* = -\infty$.

Step 2: Solving the decomposed subproblems parallelly

Step 2.1 Solving the timetable and bus formation plan subproblem $\underline{\boldsymbol{P_y}}$

For *each direction* $d \in D$ **do**

Find the shortest path from vertex $(t_0, \boldsymbol{p_0})$ to $(|T|,*)$ using DP;

Let $y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'} = 1$ for arc $(t, t', \boldsymbol{p}, \boldsymbol{p}')$ along the shortest path;

End

Step 2.2 Solving the vehicle scheduling subproblem $\underline{\boldsymbol{P_x}}$

Call the Gurobi solver to solve the vehicle scheduling subproblem $\underline{\boldsymbol{P_x}}$;

Step 3: Update the best lower bound value

Calculate the lower bound value LB^m by Equation (30) and update the best lower bound value by $LB^* = \max\{LB^*, LB^m\}$

Step 4: Update the Lagrangian multipliers

Update the value of the Lagrange multipliers $\underline{\lambda^{m+1}_{d,s^*,t}}$ in the $(m+1)th$ iteration by the following equation:

$$\underline{\lambda^{m+1}_{d,s^*,t}} = \underline{\lambda^{m}_{d,s^*,t}} + \frac{1}{m+1} \times \left(\sum_{(t,t',\boldsymbol{p},\boldsymbol{p}') \in A_d} \left(p_{s^*} \times y^d_{t,t',\boldsymbol{p},\boldsymbol{p}'} \right) - \sum_{a \in \Psi^d_{s^*,t}} \sum_{k \in K} x^k_a \right)$$

Step 5: Terminal conditions

if $m < M$, let $m = m + 1$, and go back to step 2; otherwise, output the best lower bound solution and terminate the algorithm.

References

Bertsekas, D.P., 1997. Nonlinear Programming. J. Oper. Res. Soc. 48 (3), 334.

Bhoopalam, A.K., Agatz, N., Zuidwijk, R., 2018. Planning of truck platoons: A literature review and directions for future research. Transp. Res. B Methodol. 107, 212–228.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends®*. Mach. Learn. 3 (1), 1–122.

Carosi, S., Frangioni, A., Galli, L., Girardi, L., Vallese, G., 2019. A matheuristic for integrated timetabling and vehicle scheduling. Transp. Res. B Methodol. 127, 99–124.

Ceder, A., 2016. Public transit planning and operation: Modeling, practice and behavior. CRC Press.

Ceder, A., Wilson, N.H.M., 1986. Bus network design. Transp. Res. B Methodol. 20 (4), 331–344.

Chen, C., He, B., Ye, Y., Yuan, X., 2016. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math. Program. 155 (1), 57–79.

Chen, X., He, S., Zhang, Y., Tong, L., Shang, P., Zhou, X., 2020a. Yard crane and AGV scheduling in automated container terminal: A multi-robot task allocation framework. Transportation Research Part C: Emerging Technologies 114, 241–271.

Chen, Z., 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., 2019. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Discrete modeling method. Transp. Res. B Methodol. 122, 1–19.

Chen, Z., Li, X., Zhou, X., 2020b. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Continuous modeling method. Transp. Res. B Methodol. 132, 76–100.

Chen, Z., Li, X., Qu, X., 2022. A Continuous Model for Designing Corridor Systems with Modular Autonomous Vehicles Enabling Station-wise Docking. Transp. Sci. 56 (1), 1–30.

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.
Dai, Z., Cathy Liu, X., Li, H., Wang, M., Ma, X., 2023. Semi-autonomous bus platooning service optimization with surrogate modeling. Comput. Ind. Eng. 175, 108838.
Dakic, I., Yang, K., Menendez, M., Chow, J.Y.J., 2021. On the design of an optimal flexible bus dispatching system with modular bus units: Using the three-dimensional macroscopic fundamental diagram. Transp. Res. B Methodol. 148, 38–59.
Fonseca, J.P., van der Hurk, E., Roberti, R., Larsen, A., 2018. A matheuristic for transfer synchronization through integrated timetabling and vehicle scheduling. Transp. Res. B Methodol. 109, 128–149.
Gong, M., Hu, Y., Chen, Z., Li, X., 2021. Transfer-based customized modular bus system design with passenger-route assignment optimization. Transportation Research Part E: Logistics and Transportation Review 153, 102422.
Hannoun, G.J., Menéndez, M., 2022. Modular vehicle technology for emergency medical services. Transportation Research Part C: Emerging Technologies 140, 103694.
Hassold, S., Ceder, A., 2012. Multiobjective Approach to Creating Bus Timetables with Multiple Vehicle Types. Transp. Res. Rec. 2276 (1), 56–62.
He, B.S., Yang, H., Wang, S.L., 2000. Alternating Direction Method with Self-Adaptive Penalty Parameters for Monotone Variational Inequalities. J. Optim. Theory Appl. 106 (2), 337–356.
Hu, B., Fu, Y., Feng, S., 2023. Integrated optimization of multi-vehicle-type timetabling and scheduling to accommodate periodic passenger flow. Comput. Aided Civ. Inf. Eng. 1–25.
Ji, Y., Liu, B., Shen, Y., Du, Y., 2021. Scheduling strategy for transit routes with modular autonomous vehicles. Int. J. Transp. Sci. Technol. 10 (2), 121–135.
Khan, Z.S., He, W., Menéndez, M., 2023. Application of modular vehicle technology to mitigate bus bunching. Transportation Research Part C: Emerging Technologies 146, 103953.
Kliewer, N., Mellouli, T., Suhl, L., 2006. A time–space network based exact optimization model for multi-depot bus scheduling. Eur. J. Oper. Res. 175 (3), 1616–1627.
Liu, T., Ceder, A., 2018. Integrated public transport timetable synchronization and vehicle scheduling with demand assignment: A bi-objective bi-level model using deficit function approach. Transp. Res. B Methodol. 117, 935–955.
Liu, T., Ceder, A., Rau, A., 2020. Using Deficit Function to Determine the Minimum Fleet Size of an Autonomous Modular Public Transit System. Transp. Res. Rec. 2674 (11), 532–541.
Liu, X., Qu, X., Ma, X., 2021a. Improving flex-route transit services with modular autonomous vehicles. Transportation Research Part E: Logistics and Transportation Review 149, 102331.
Liu, X., Qu, X., Ma, X., 2021b. Optimizing electric bus charging infrastructure considering power matching and seasonality. Transp. Res. Part D: Transp. Environ. 100, 103057.
Liu, X., Liu, X., Zhang, X., Zhou, Y., Chen, J., Ma, X., 2023. Optimal location planning of electric bus charging stations with integrated photovoltaic and energy storage system. Comput. Aided Civ. Inf. Eng. 38 (11), 1424–1446.
Niu, H., Zhou, X., 2013. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. Transportation Research Part C: Emerging Technologies 36, 212–230.
Repoux, M., Geroliminis, N., Kaspi, M., 2021. Operational analysis of an innovative semi-autonomous on-demand transportation system. Transportation Research Part C: Emerging Technologies 132, 103373.
Scherr, Y.O., Neumann Saavedra, B.A., Hewitt, M., Mattfeld, D.C., 2019. Service network design with mixed autonomous fleets. Transportation Research Part E: Logistics and Transportation Review 124, 40–55.
Scherr, Y.O., Hewitt, M., Mattfeld, D.C., 2022. Stochastic service network design for a platooning service provider. Transportation Research Part C: Emerging Technologies 144, 103912.
Shafahi, Y., Khani, A., 2010. A practical model for transfer optimization in a transit network: Model formulations and solutions. Transp. Res. A Policy Pract. 44 (6), 377–389.
Shang, P., Li, R., Liu, Z., Yang, L., Wang, Y., 2018. Equity-oriented skip-stopping schedule optimization in an oversaturated urban rail transit network. Transportation Research Part C: Emerging Technologies 89, 321–343.
Shi, X., 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.
Stevens, M., Correia, G.H.d.A., Scheltes, A., van Arem, B., 2022. An agent-based model for assessing the financial viability of autonomous mobility on-demand systems used as first and last-mile of public transport trips: A case-study in Rotterdam, the Netherlands. Res. Transp. Bus. Manag. 45, 100875.
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., 2023. Joint scheduling and formation design for modular-vehicle transit service with time-dependent demand. Transportation Research Part C: Emerging Technologies 147, 103986.
Völz, V., Mattfeld, D.C., Scherr, Y.O., 2022. Relocation planning with partly autonomous vehicles in car sharing systems. Transp. Res. Procedia 62, 213–220.
Wang, S., Correia, G.H.d.A., Lin, H.X., Severino, A., 2022. Assessing the Potential of the Strategic Formation of Urban Platoons for Shared Automated Vehicle Fleets. J. Adv. Transp. 2022, 1–20.
Wu, J., Kulcsár, B., Selpi, Qu, X., 2021. A modular, adaptive, and autonomous transit system (MAATS): An in-motion transfer strategy and performance evaluation in urban grid transit networks. *Transportation Research Part A: Policy and Practice* 151, 81-98.
Wu, M., Yu, C., Ma, W., An, K., Zhong, Z., 2022. Joint optimization of timetabling, vehicle scheduling, and ride-matching in a flexible multi-type shuttle bus system. Transportation Research Part C: Emerging Technologies 139, 103657.
Yang, J., Jin, J.G., Wu, J., Jiang, X., 2017. Optimizing Passenger Flow Control and Bus-Bridging Service for Commuting Metro Lines. Comput. Aided Civ. Inf. Eng. 32 (6), 458–473.
Yang, L., Yao, Y., Shi, H., Shang, P., 2021. Dynamic passenger demand-oriented train scheduling optimization considering flexible short-turning strategy. J. Oper. Res. Soc. 72 (8), 1707–1725.
Yao, Y., Zhu, X., Dong, H., Wu, S., Wu, H., Carol Tong, L., Zhou, X., 2019. ADMM-based problem decomposition scheme for vehicle routing problem with time windows. Transp. Res. B Methodol. 129, 156–174.
Yu, B., Yang, Z., Sun, X., Yao, B., Zeng, Q., Jeppesen, E., 2011. Parallel genetic algorithm in bus route headway optimization. Appl. Soft Comput. 11 (8), 5081–5091.
Yuan, J., Gao, Y., Li, S., Liu, P., Yang, L., 2022. Integrated optimization of train timetable, rolling stock assignment and short-turning strategy for a metro line. Eur. J. Oper. Res. 301 (3), 855–874.
Zhan, S., Wong, S.C., Shang, P., Peng, Q., Xie, J., Lo, S.M., 2021. Integrated railway timetable rescheduling and dynamic passenger routing during a complete blockage. Transp. Res. B Methodol. 143, 86–123.
Zhang, Y., Peng, Q., Yao, Y., Zhang, X., Zhou, X., 2019. Solving cyclic train timetabling problem through model reformulation: Extended time-space network construct and Alternating Direction Method of Multipliers methods. Transp. Res. B Methodol. 128, 344–379.
Zhang, Y., Peng, Q., Lu, G., Zhong, Q., Yan, X., Zhou, X., 2022. Integrated line planning and train timetabling through price-based cross-resolution feedback mechanism. Transp. Res. B Methodol. 155, 240–277.
Zhuo, S., Miao, J., Meng, L., Yang, L., Shang, P., 2023. Demand-driven integrated train timetabling and rolling stock scheduling on urban rail transit line. Transport Science, Transportmetrica A, pp. 1–42.