Operations Design of Modular Vehicles on an Oversaturated Corridor with First-in-first-out Passenger Queueing

Xiaowei Shi, Xiaopeng Li*

Department of Civil and Environmental Engineering, University of South Florida, 4202 E Fowler Avenue, ENC 3300, Tampa, FL 33620, United States

Contact: xiaoweishi@mail.usf.edu, xiaopengli@usf.edu

Abstract: While urban transit systems (UTS) often have fixed vehicle capacity and relatively constant departure headways, they may need to accommodate dramatically fluctuating passenger demands over space and time, resulting in either excessive passenger waiting or vehicle capacity and energy waste. Therefore, on the one hand, optimal operations of UTS rely on accurate modeling of passenger queuing dynamics, which is particularly complex on a multi-stop transit corridor. On the other hand, capacities of transit vehicles can be made variable and adaptive to time-variant passenger demand so as to minimize energy waste, especially with the emergence of Modular Vehicle (MV) technologies. This paper investigates operations of a multi-stop transit corridor where vehicles may have different capacities across dispatches. We specify skewed time coordinates to simplify the problem structure while incorporating traffic congestion. Then we propose a mixed integer linear programming model that determines the optimal dynamic headways and vehicle capacities over study time horizon to minimize the overall system cost for the transit corridor. In particular, the proposed model considers a realistic Multi-Stop-First-In-First-Out (MSFIFO) rule that gives the same boarding priority to passengers arriving at a station in the same time interval yet with different destinations. A customized Dynamic Programming (DP) algorithm is proposed to solve this model efficiently. To circumvent the rapid increase of the state space of a typical DP algorithm, we analyze the theoretical properties of the investigated problem and identify upper and lower bounds to a feasible solution. The bounds largely reduce the state space during the DP iterations and greatly improve the efficiency of the proposed DP algorithm. The state dimensions are also reduced with the MSFIFO rule such that all queues with different destinations at the same origin can be tracked with a single boarding position state variable at each stage. A hypothetical example and a real-world case study show that the proposed DP algorithm greatly outperforms a state-of-the-art commercial solver (Gurobi) in both solution quality and time.

Key words: Modular vehicle; variable-capacity operation; schedule; overall system cost; dynamic programming.

1. Introduction

Despite facing dramatic fluctuations of passenger demand over both time and space, urban transit systems (UTS) often feature relatively constant departure headways and fixed transit vehicle capacity. This demand-supply mismatch may cause either excessively long customer waiting time (e.g. in peak-hours or densely populated areas) or waste of vehicle capacity and energy (e.g. in off-peak hours or areas with sparse demand), which often results in deteriorating service quality and increased safety risk and operating cost (Li and Lo, 2014; Gao et al., 2016; Yang et al., 2016; Cacchiani et al., 2016; Shi et al., 2018). A report conducted by the New York UTS (https://toddwschneider.com) shows that the average waiting time for passengers in a day varies widely from 2 to 30 minutes.

A classical method of reducing passenger waiting time is the dynamic adjustment of the vehicle dispatch headway to better accommodate demand variations, e.g., a smaller headway for higher demands

(Barrena et al., 2014a; Canca et al., 2014; Huang et al., 2016; Niu et al., 2015). Yet due to limited resources (e.g., limited vehicle capacity) and minimum safety headway constraints, this method may not completely eliminate passenger queues in peak hours. Passenger demand greatly exceeding UTS capacity supply may result in long passenger queues, and some passengers may have to wait for several trains to pass before being able to board. The phenomenon of a passenger queue that exceeds the capacity of the passing vehicles, compelling some passengers to wait for later vehicles, is called oversaturated queuing (Niu and Zhou, 2013; Zhao et al., 2018; Chen et al., 2019). Dynamics of the Oversaturated Queuing Phenomenon (OQP) poses significant challenges to UTS modeling. In particular, when the OQP is present in a multi-stop transit corridor where passengers in the same queue at a station may have different destinations, it is quite challenging to model the boarding orders of passengers with different destinations.

Only a handful of studies address the OQP on a transit corridor with multiple destinations (e.g. Niu and Zhou, 2013; Yin et al., 2016; Shang et al., 2018; Shi et al., 2018). Despite these successes, it is yet inconclusive how to both efficiently and realistically model actual passenger boarding orders in practice, i.e. passengers arriving earlier than excepted to board sooner, or the first-in-first-out queuing rule (FIFO). To ensure the realistic FIFO rule, some studies (e.g., Niu and Zhou, 2013; Shang et al., 2018) divide the entire time horizons into time intervals such that no more than one passenger arrives at a station in the same time interval. This single-arrival setting eliminates the difficulty of assigning the boarding portions of passengers with different destinations in a time interval (since only one passenger with one destination needs to be assigned at a time). However, when the demand is high (e.g. peak hours), many passengers may arrive at a station in a rather short period. Then the validity of the single-arrival setting requires the time horizon to be divided into an extremely large number of small intervals to ensure no more than one passenger arrives in one interval. This would dramatically increase the problem size and associated complexity and render the model computationally intractable for realistic large-scale instances.

However, the boarding priorities of passengers with different destinations arriving in the same interval become a modeling challenge. One way to circumvent this challenge is to assume that even if an oncoming vehicle has vacant seats but cannot accommodate all waiting passengers that arrived in one time interval, none of the passengers is allowed to board the vehicle (e.g. Yin et al., 2016). While this All-Or-Nothing Boarding Setting (AONBS) largely simplifies the problem structure and enables the development of computationally-tractable models, it cannot incorporate the realistic case where waiting passengers continue to board the vehicle until all vacant seats are filled. Figure 1 illustrates the problem of the all-or-nothing setting with a simple three-station example. In this example, among 100 passengers arriving at Station 1 in the first time interval, 50 are destined for Station 2 and the other 50 for Station 3. In this first time interval, suppose that the two groups of passengers with different destinations are all in the same waiting queue, then the realistic FIFO rule yields that the numbers (or the portions) of the two passenger groups boarding the same vehicle are always identical. Further, all 50 passengers arriving at Station 2 in time interval 1 are destined for Station 3. Suppose a vehicle with the capacity of 50 arrives at Station 1 at the end of time interval 1, and then at Station 2 at the end of time interval 2. The boarding result from the realistic FIFO rule is shown on the left in Fig. 1. We see that the boarding numbers from the two passenger groups are identical: 25 passengers from each group boards the vehicle in time interval 2, which fills up the vehicle capacity and leaves 25 passengers of each group remaining in the queue. After the vehicle arrives at Station 2, the 25 passengers from Station 1 destined for Station 2 alight, making 25 vacant seats, and then 25 passengers from the queue at Station 2 board the vehicle, leaving 25 remaining in the queue. Overall, a total of 75 passengers complete the trip with this vehicle. However, as shown on the right in Fig. 1, with the AONBS (as opposed to the realistic FIFO rule), since the capacity of the vehicle cannot accommodate all 100 passengers arriving in the same time interval waiting at Station 1, no passenger is allowed to board the vehicle. As a result, an empty vehicle arrives at Station 2 and takes away all 50 passengers waiting at Station 2. This is apparently unrealistic.

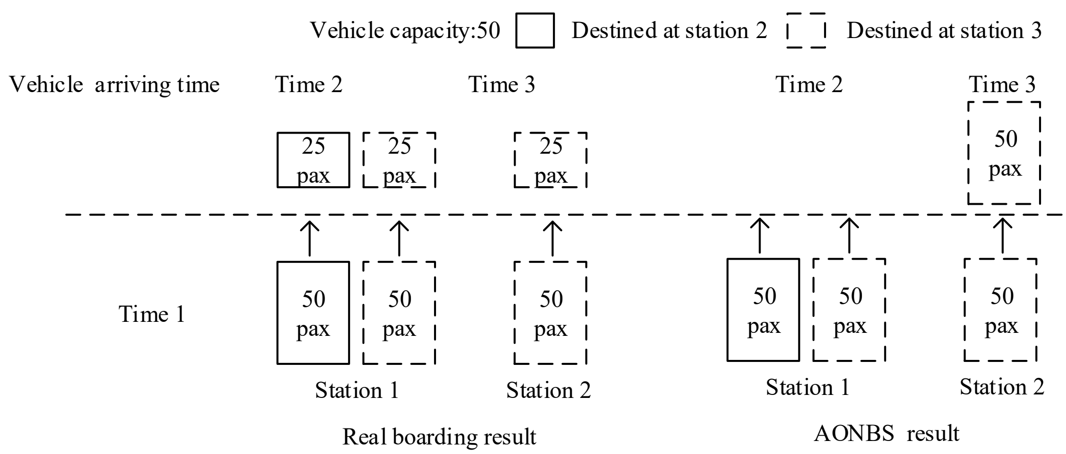


Fig. 1 An instance to illustrate the boarding result of the AONBS (pax is short for passengers).

To avoid the vacant seat problem of the AONBS, other studies (e.g., Shi et al., 2018; Zhao et al., 2018) let the model optimize the boarding priorities of the passengers arriving in the same interval to minimize the number of the dispatched vehicles (e.g., via maximizing the number of cumulative boarding passengers served per vehicle). While this Priority Boarding Setting (PBS) ensures that all passengers will be seated whenever the vehicle capacity allows, it may lead to a boarding order quite different from the actual FIFO rule. Note that in the FIFO rule, passengers are not really coordinated in their boarding orders, and whoever arrives at the station earlier should board the vehicle sooner regardless of their destinations. For example, Fig. 2 illustrates the problem of the PBS with the same instance as Fig. 1, where again passengers with different destinations arrive simultaneously at a station in each time interval. To minimize the number of the dispatched vehicles (or maximize the number of cumulative boarding passengers), it is observed on the right in Fig. 2 that only passengers destined for Station 2 is allowed to board the vehicle at Station 1, resulting in a total number of cumulative boarding passengers of 100, which is greater than the boarding result (i.e. 75) from the actual FIFO rule.

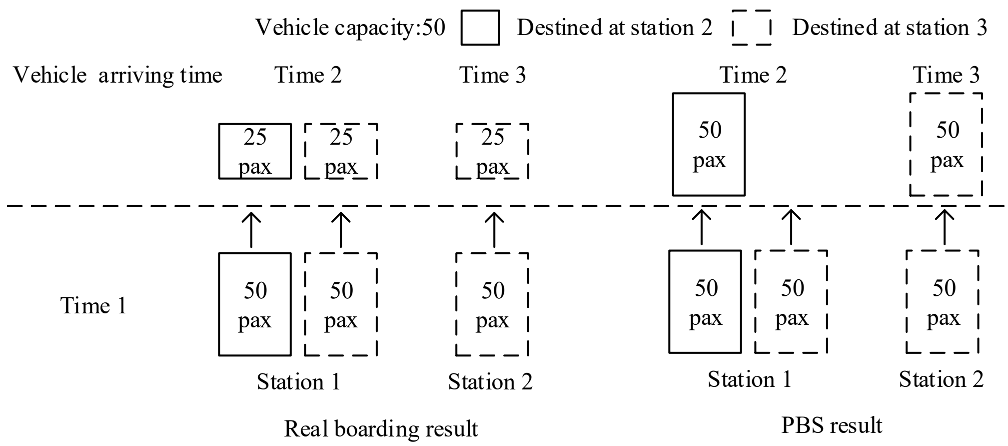


Fig. 2 An instance to illustrate the boarding result of the PBS.

Most studies addressing the OQP only consider fixed-capacity vehicles since many traditional transit vehicle technologies do not allow flexible changes of vehicle capacity during operations. Yet this fixed capacity constraint may be relaxed due to the emergence of Modular Vehicle (MV) technology. With this technology, a vehicle can be composed of an indefinite number of Modular Units (MU) and thus can flexibly vary its capacity by concatenating or detaching MUs across different dispatches. This MV concept of variable-capacity operations has been investigated in multiple transportation modes, including transit bus scheduling (Ceder, 2011; Hassold and Ceder, 2014), suburban railway scheduling (Albrecht, 2009), transit switching (Guo et al., 2018; Caros, 2019), and shuttle design (Chen et al., 2019a, 2019b). Please see Liu and Ceder (2017) for a recent review on this topic. The results from these studies suggest the potential of the variable-capacity operations on reducing overall system cost and increasing service quality. However, as far as we know, there is still a shortage of studies investigating the MV concept in the corridor transit system, in particular when the OQP is present. The most relevant study to our work is Mo et al. (2019). It incorporates the MV concept in an urban rail transit corridor and proposes a nonlinear integer programming model to obtain the optimal train schedule with minimum system cost. However, the paper does not consider the OQP, and the quality of the solution obtained by a tabu-search-based algorithm can be further improved.

In addition, due to variations in traffic pattern, travel time may also vary across different dispatch time. In the literature, Barrena et al. (2014b) and Gao et al. (2016) consider that travel time may be impacted by dynamic passenger demand patterns. Niu and Zhou (2013) consider time-dependent travel time due to periodic traffic congestion (i.e. travel time is only determined by the vehicle dispatch time). However, time-dependent travel time, albeit relevant, is not addressed in the OQP context while incorporating the MV concept.

To fill these research gaps, this paper develops an operation design model for a transit corridor with variable-capacity MVs considering the OQP. The investigated problem is presented in time coordinates skewed for each station to offset the MV travel time, which greatly simplifies the problem structure while incorporating time-dependent travel time due to traffic congestion. We first tackle the challenge of formulating the realistic FIFO boarding orders under the OQP by devising a computational, efficient Multi-Stop-FIFO (MSFIFO) rule. The MSFIFO rule assumes that passengers with different destinations arrive simultaneously at a station in each time interval and treats the passengers with the same destination arriving at the station in the same time interval as a passenger group. This arrival pattern imposes that if a vehicle's capacity cannot accommodate all passengers arriving in one-time interval at a station, the number of boarding passengers from each group is proportional to the group size. The proposed MSFIFO rule specifies a set of auxiliary continuous variables and an associated queuing structure to overcome the challenge of tracking the boarding portion of each group as time elapses. This innovation not only correctly tracks the passenger boarding counts but also reduces the problem dimensionality from dealing with passengers individually (Niu and Zhou, 2013; Shang et al., 2018) to processing them in batches, which consequently improves the solution efficiency.

The proposed MSFIFO rule enables the formulation of a compact Mixed Integer Linear Programming (MILP) model for the joint design of optimal dynamic headways and vehicle capacities under the OQP. The optimization objective includes passenger waiting cost and MV operating cost as a trade-off, subject to constraints on vehicle dispatch, vehicle capacity, and the MSFIFO rule. Due to the complex nature of the investigated problem, despite the compact form of the proposed model, state-of-the-art commercial solvers (e.g. Gurobi) still cannot solve the large-scale real-world instances. To overcome this limit, this paper studies the theoretical properties of the investigated problem and identifies upper and lower bounds to a feasible solution. The theoretical findings lead to the development of a customized efficient Dynamic Programming (DP) algorithm with a reduced state space. The state dimensions are also reduced with the MSFIFO rule such that all queues with different destinations at the same origin can be tracked with a single boarding position state variable at each stage. Numerical results show that the customized DP algorithm greatly outperforms a state-of-the-art commercial solver (Gurobi). Further, they reveal interesting insights into the advantages of the joint design of dynamic headway and variable vehicle capacity over existing operation paradigms. The developments in this paper will serve as a theoretical and methodological basis for UTS operations, particularly for near-future systems with emerging MV technology.

The rest of this paper is organized as follows. Section 2 states the investigated problem of MV Operations Design on an Oversaturated Corridor (MODOC) and formulates it into a MILP model with the MSFIFO rule. Section 3 proposes a customized DP algorithm to solve the proposed model efficiently. Section 4 conducts two case studies (including an illustrative hypothetical case and a large-scale real-world case based on the operation data from Beijing Subway Line 6) to demonstrate the efficiency of the proposed solution approach and draw managerial insights. Section 5 briefly discusses conclusions and further research directions.

2. Mathematical formulation

2.1 Notation

For readers' convenience, the notion in this paper is summarized in Tab. 1.

Tab. 1 Notation list.

Parameters	
$\mathcal{I}$	Set of stations, $\mathcal{I} = \{1,2,\cdots,I\}$.
$\mathcal{I}_i^+$	Set of stations downstream of Station i, $\mathcal{I}_i^+ = \{i+1, i+2, \cdots, I+1\}$ $\forall i \in \mathcal{I}$. Station $I+1$ is a dummy station without arriving and departure passengers.
$\mathcal{I}_i^-$	Set of stations upstream of Station i, $\mathcal{I}_i^- = \{0,1,2,\cdots,i-1\}$ $\forall i \in \mathcal{I}$. Station 0 is a dummy station without arriving and departure passengers.
$\mathcal{T}$	Set of time intervals, $\mathcal{T} = \{1,2,\cdots,T\}$.
$\mathcal{L}$	Set of MV capacity levels $\mathcal{L} = \{1,2,\cdots,L\}$; $l \in \mathcal{L}$ also denotes the number of MUs of the MV.
δ	Duration of a time interval (e.g. 1 min, 2 min, 5min).
$p_{ijt'}$	Size of passenger group ijt', i.e., the group of passengers arriving at Station i at time interval t' destined for Station j, $\forall i \in \mathcal{I}, j \in \mathcal{I}_i^+, t' \in \mathcal{T}$.
C	Capacity of one single MU.
V	Total number of MUs in stock.
H	Minimum headway.
P	Operational cycle.
$e(l)$	Operating cost consumed in dispatching a level-l MV, $\forall l \in \mathcal{L}$.
w	Value of waiting time for each passenger.
g_t	MU pipeline stock state in dynamic programming.
o_t	MU terminal stock state in dynamic programming.
v_t	Boarding position state in dynamic programming.
M_t	Total cost till Stage t in dynamic programming.
Decision variables	
y_{lt}	=1, if a level-l MV (with l MUs) is dispatched at the end of time interval t; =0, otherwise, $\forall l \in \mathcal{L}, t \in \mathcal{T}$.
c_{it}	Amount of vacant seat of the MV dispatched at the end of time interval t on arriving at Station i, $\forall i \in \mathcal{I}, t \in \mathcal{T}$.
$x_{it't}$	=1, if all passengers in group it' board the MV till the end of time interval t; =0, otherwise, $\forall i \in \mathcal{I}, t' \in \mathcal{T}, t \geq t' \in \mathcal{T}$. We omit index j since this percentage is the same across all passenger groups arriving at Station i in time interval t' due to the MSFIFO rule.
$u_{it't}$	Cumulative boarding percentage of passenger group ijt' till the end of time interval t, $\forall i \in \mathcal{I}, j \in \mathcal{I}_i^+, t' \in \mathcal{T}, t \geq t' \in \mathcal{T}$.
Y_t	Decision variable in dynamic programming.

2.2 Problem statement

The investigated problem of MV Operations Design on an Oversaturated Corridor (MODOC) is illustrated in Fig. 3. The MV UTS corridor contains I stations denoted as set $\mathcal{I} = \{1,2,\cdots,I\}$, where the index increases from upstream to downstream. Let $[0, D]$ denote the study time horizon. To facilitate the model construction, we divide the time horizon equally into T fixed-length time intervals $\mathcal{T} = \{1,2,\cdots,T\}$, each with duration δ. The set of MV levels is $\mathcal{L} = \{1,2,\cdots,L\}$, where $l \in \mathcal{L}$ also denotes the number of MUs of the corresponding MV. The capacity of a single MU is C, and thus a level-l MV has capacity of lC. Let $e(l)$ denote the operating cost of dispatching a level-l MV. By referring to Chen et al. (2019a), we find that the operating cost function $e(l)$ is concave over the MV level l, and it can be expressed by $e(l) = \lambda^F + \lambda^V (lC)^\alpha \; \forall l \in \mathcal{L}$, where λ^F is the fixed energy cost, λ^V is a positive coefficient related to the number of dispatched MUs, and power index $\alpha \leq 1$. In general, the operating cost of dispatching a vehicle with C seats is far less than the passenger waiting cost arising from the corresponding number of passengers (i.e. C) waiting for one time interval. Since the operating cost function is concave, the relationship between the operating cost and passenger waiting cost is expressed by $e(1) \ll w\delta C$, where w denotes the value of waiting time for each passenger. There are V MUs in stock circulating on the corridor. Each MV takes P time intervals to complete an operational cycle from being dispatched from Station 1 to circulating back to Station 1. After each operational cycle, MUs in an MV can be immediately re-organized into other MVs of different lengths for the following dispatches. The minimum dispatch headway between any two consecutive vehicles is H time intervals. Fig. 3 provides a toy instance of the system. It investigates a UTS with 3 stations in a time horizon divided into 4 time intervals (i.e. I=3, T=4). The maximum capacity level is 6 and the capacity of a single MU is 5 passengers (i.e. L=6, C=5). There are 10 MUs in stock (i.e. V=10). The operational cycle and minimum dispatch headway are 2 and 1 time intervals respectively (i.e. P=2,H=1). The operating cost of dispatching a level-l MV is l.

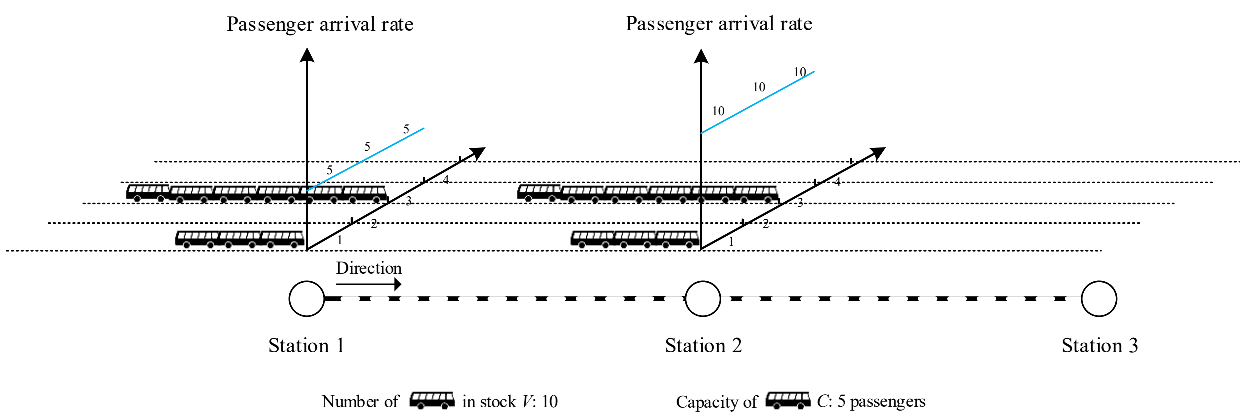


Fig. 3 Problem illustration.

We assume that the travel time (including the dwell time) of an MV between any two consecutive stations only depends on the MV's dispatch time, regardless of vehicle capacity levels (since capacity does not affect travel speed) or passenger boarding and aligning amounts (assuming the dwell time at each station is fixed), and no overtaking is allowed in the operation. Motivated by Sun et al. (2014), we skew each station's time coordinates, such that in the skewed coordinates the time when an MV arrives at the station is identical to the MV's dispatch time at Station 1. This time coordinates skewing operation is described in Fig. 4. We see that vehicles 1 and 2 are dispatched from Station 1 at time interval 1 and time interval 3, respectively. The travel time for vehicle 1 from Station 1 to 2 is one time interval, and that from Station 2 to 3 is one time interval as well. The travel time for vehicle 2 from Station 1 to 2 is three time intervals, and that from Station 2 to 3 is one time interval, as shown on the left in Fig. 4. With that, we can skew the time coordinates of Stations 2 and 3, and make the arriving time of vehicles 1 and 2 at each station consistent with its dispatch time, as shown on the right in Fig. 4. Now, vehicles 1 and 2 have the same arriving time at each station during its operation process with the skewed time coordinates. To specify it, define a mapping function $K_i(t)$ $\forall i \in \mathcal{I}, t \in \mathcal{T}$ to denote the original time point before the skewing operation corresponding to time t in the skewed time coordinates at Station i. In this instance (i.e., Fig. 4), $K_1(1)$=1, $K_1(2)$=2, $K_1(3)$=3; $K_2(1)$=2, $K_2(2)$=4, $K_2(3)$=6; $K_3(1)$=3, $K_3(2)$=5, $K_3(3)$=7.

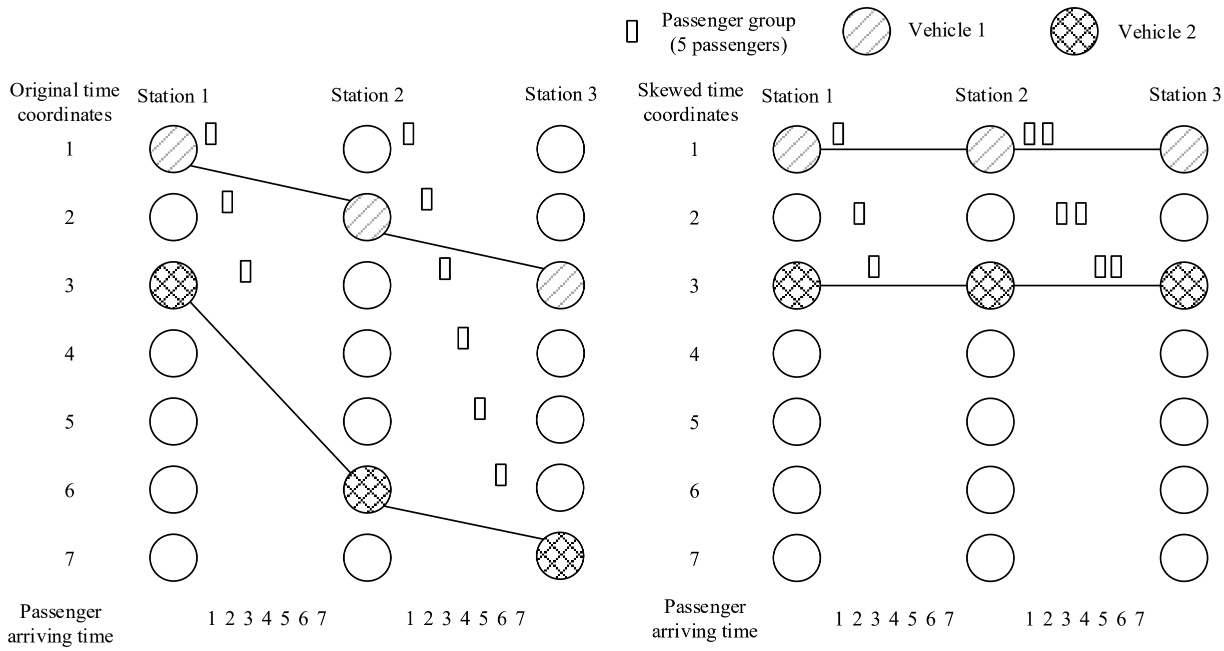


Fig. 4 The process for arriving time revision.

Note that the time coordinates skewing operation may also skew the arrival demand. In the original time coordinates, let $p^0_{ijt'}$ denote the size of passenger group ijt', i.e. the group of passengers arriving at Station i in original time interval t' destined for Station j. If $t' > T$, we denote $p^0_{ijt'}$=0; if t'=0, p^0_{ij0}=0. Then, in the skewed time coordinates, the corresponding size of passenger group $p_{ijt'}$can be obtained by $p_{ijt'} := \sum_{K_i(t'-1)}^{K_i(t')} p^0_{ijt'}$ $\forall t' \in \mathcal{T}$, i.e. the group of passengers arriving at Station i in skewed time interval t' destined for Station j. As illustrated in Fig. 4, squares represent identical passenger groups, and each group includes 5 passengers. In the original time coordinates, one passenger group arrives at each station in one time interval, as shown on the left in Fig. 4. In the skewed time coordinates, at Station 1, the arrival demand is the same as the arrival demand in the original time coordinates; at Station 2, according to the proposed equation, the size of the passenger groups arriving at time 1 equals 10 passengers (i.e. ($\sum_{i,j} p_{ij0}$+$\sum_{i,j} p_{ij1}$+$\sum_{i,j} p_{ij2}$)·5), the size of the passenger groups arriving at time 2 equals 10 passengers (i.e. ($\sum_{i,j} p_{ij3}$+$\sum_{i,j} p_{ij4}$) ·5), and the size of the passenger groups arriving at time 3 equals 10 passengers (i.e., ($\sum_{i,j} p_{ij5}$+$\sum_{i,j} p_{ij6}$) ·5).

This time coordination conversion greatly simplifies the notation. In the skewed time coordinates, note that the same MV will arrive at each station at the same time. Thus, we can just use capacity level and dispatch time to index an MV, i.e. MV lt denoting the MV with the level-l capacity dispatched from station

1 at the end of time interval t (or at time t for short). The operations design decisions include the time and level of each MV dispatch, which is denoted by the binary variable $y_{lt} = 1$ if a level-l MV is dispatched at time t. With the skewed time coordinates, the proposed model no longer needs to track a vehicle's arrival time at each single station as required in the literature (e.g., Barrena et al., 2014b; Cacchiani et al., 2016; Gao et al., 2016; Yin et al., 2017), and this greatly simplifies the model structure. Further, it incorporates time-dependent travel time in the presence of periodic traffic congestion, which is seldom investigated in the related transit literature.

We assume that in each time interval, each passenger group's arrival rate is constant, and thus the passenger arrival, queueing and boarding behavior follows the MSFIFO rule. With this MSFIFO rule, if a vehicle's vacant seats cannot accommodate all passengers arriving in one time interval at a station, the number of boarding passengers from each group is proportional to the group size. As a result, the boarding percentages of different passenger groups arriving at the same station in the same time interval are all identical. In a case where passengers cannot board immediately upon arrival, again w denotes the value of waiting time for each passenger. In the toy instance, we set w to 1.

The MODOC problem is to find the optimal MV dispatch schedules and the corresponding capacity levels in the study time horizon to balance the trade-off between the passenger waiting cost and vehicle operating cost. As shown in Fig. 3, for this instance, a level-3 MV is dispatched at the end of time interval 1, and a level-6 MV is dispatched at the end of time interval 3 to minimize the passenger waiting cost (i.e., 85) and vehicle operating cost (i.e., 9). The operation process can be described as follows. At the end of time interval 1, a level-3 MV is dispatched and the number of MUs in stock becomes 7 (i.e., V=7). At Station 1, 5 passengers board the level-3 MV; at Station 2, 10 passengers board the level-3 MV; all passengers alight at Station 3. At the end of time interval 2, due to the minimum dispatch headway, no MV is dispatched thus 5 and 10 passengers arriving in time interval 2 have to wait at Stations 1 and 2, respectively. At the end of time interval 3, the level-6 MV is dispatched and the number of MUs in stock becomes 1 (i.e., V=1). At Station 1, 5 passengers arriving at time interval 2 and 5 passengers arriving at time interval 3 board the level-6 MV; at Station 2, 10 passengers arriving at time interval 2 and 10 passengers arriving at time interval 3 board the level-6 MV; and all passengers alight at Station 3. At the end of time interval 4, no MV is dispatched due to the minimum dispatch headway, and 3 MUs dispatched at time interval 1 go back to the stock (i.e., V=4).

2.3 Model Formulation

This section formulates the studied problem into a MLIP model. The model includes three sets of constraints on the vehicle dispatch, vehicle capacity and MSFIFO rule, respectively, and an objective function including passenger waiting cost and vehicle operating cost.

2.3.1 Vehicle dispatch constraints

The set of constraints related to vehicle dispatches are formulated as Equations (1)-(3). Constraint (1) indicates that in an operational cycle P, the total number of the dispatched MUs cannot exceed the number of MUs in stock. Constraint (2) guarantees two consecutive MV dispatches are separated at least by a safety time headway of H intervals. Constraint (3) sets the binary domain of vehicle dispatch variables (noting that y_{lt}=1 if a level-l MV (with l MUs) is dispatched at time, or y_{lt}=0 otherwise).

$$\sum_{t=t''}^{t''+P} \sum_{l \in \mathcal{L}} l y_{lt} \leq V, \forall t'' \in \{1,2,\cdots,T-P\}, \tag{1}$$

$$\sum_{t=t''}^{t''+H} \sum_{l \in \mathcal{L}} y_{lt} \leq 1, \forall t'' \in \{1,2,\cdots,T-H\}, \tag{2}$$

$$y_{lt} \in \{0,1\}, \forall l \in \mathcal{L}, t \in \mathcal{T}. \tag{3}$$

2.3.2 Vehicle capacity constraints

The set of constraints related to vehicle capacity are formulated as Equations (4)-(7). To clearly describe passenger boarding and alighting dynamics, we define two sets of auxiliary decision variables, i.e., vacant seat variables $\{c_{it} \in \mathbb{R}^+, \forall i \in \mathcal{I}, t \in \mathcal{T}\}$ and cumulative boarding percentage variables $\{u_{it't} \in [0,1], \forall i \in \mathcal{I}, t' \in \mathcal{T}, t \geq t' \in \mathcal{T}\}$. Variable c_{it} represents the amount of vacant seats of the MV dispatched at the end of time interval t on arriving at Station i before passengers boarding or alighting. We consider $c_{it} \gg 1$ for most of the time, and thus defining it as a continuous variable is reasonable from a macroscopic perspective. Variable $u_{it't}$ indicates the cumulative boarding percentage of passenger group ijt' (i.e., the group of passengers arriving at Station i in time interval t' destined for Station j) till the end of time interval t (equal to or greater than their arrival time t').

Constraint (4) describes the dynamics of vacant seats of a MV during two consecutive stations, i.e., the number of vacant seats at Station $i+1$ equals that of the same MV at Station i plus the alighting amount and minus the boarding amount. For simplification purpose, we denote upstream and downstream stations of Station i by $\mathcal{I}_i^- = \{0,1,2,\cdots,i-1\}$ and $\mathcal{I}_i^+ = \{i+1,i+2,\cdots,I+1\}, \forall i \in \mathcal{I}$. Station 0 and Station $I+1$ are dummy stations thus the number of arriving and departure passengers for these two stations is 0 during the study time horizon (i.e., $p_{0jt} = 0, \forall j \in \mathcal{I}, t \in \mathcal{T}$, $p_{i(I+1)t} = 0, \forall i \in \mathcal{I}, t \in \mathcal{T}$). Constraint (5) sets the initial number of vacant seats of an MV before arriving at Station 1 to its corresponding capacity. Constraint (6) postulates that the number of vacant seats throughout all stations are non-negative to ensure that the MV capacity is not exceeded. Constraint (7) indicates the range of cumulative boarding percentage of passenger group ijt' till the end of time interval t, which is between 0 and 1 (or 100%).

$$c_{(i+1)t} = c_{it} + \sum_{j \in \mathcal{I}_i^-, t' \leq t \in \mathcal{T}} \left(u_{jt't} - u_{jt'(t-1)}\right) p_{jit'} - \sum_{j \in \mathcal{I}_i^+, t' \leq t \in \mathcal{T}} \left(u_{it't} - u_{it'(t-1)}\right) p_{ijt'}, \forall i \in \mathcal{I} \backslash \{I\}, t \in \mathcal{T} \backslash \{1\}, \tag{4}$$

$$c_{1t} = \sum_{l \in \mathcal{L}} l C y_{lt}, \forall t \in \mathcal{T}, \tag{5}$$

$$0 \leq c_{it} \leq LC, \forall i \in \mathcal{I}, t \in \mathcal{T}, \tag{6}$$

$$u_{it't} \in [0,1], \forall i \in \mathcal{I}, t' \in \mathcal{T}, t \geq t' \in \mathcal{T}. \tag{7}$$

2.3.3 MSFIFO rule constraints

The set of constraints related to the MSFIFO rule are formulated as Equations (8)-(13). To address the MSFIFO rule, two sets of auxiliary decision variables, i.e., cumulative boarding percentage variables $\{u_{it't} \in [0,1]\ \forall i \in \mathcal{I}, t' \in \mathcal{T}, t \geq t' \in \mathcal{T}\}$ and boarding state variables $\{x_{it't} \in \mathbb{B}\ \forall i \in \mathcal{I}, t' \in \mathcal{T}, t \geq t' \in \mathcal{T}\}$ are defined. Binary variable $x_{it't}$ indicates the boarding state of passenger group ijt' till the end of time interval t. That is, $x_{it't}$=1 if all passengers in group ijt' board the MV till the end of time interval t, or $x_{it't}$=0 otherwise.

Constraint (8) indicates that passenger group ijt'cannot board the MV earlier than group $ij(t'-1)$. Constraint (9) imposes that if the cumulative boarding percentage of passenger group ijt' till the end of time interval t does not equal 1, variable $x_{it't}$ does not equal 1 as well. Constraint (10) means that the cumulative boarding percentage should be non-decreasing. Constraint (11) is proposed to guarantee that passengers at upstream stations have higher boarding priority than those at downstream stations. Without this constraint, to minimize the number of dispatched MVs (or maximize the number of cumulative boarding passengers), the model may force passengers at upstream stations to yield to passengers at downstream stations, which is however inconsistent with practice. Constraint (12) supposes that all passenger demand must be satisfied at the end of the time horizon and constraint (13) indicates variable $x_{it't}$ is a binary variable.

$$u_{it't} \leq x_{i(t'-1)t}, \forall i \in \mathcal{I}, t' \in \mathcal{T}\backslash\{1\}, t \geq t' \in \mathcal{T}, \tag{8}$$

$$x_{it't} \leq u_{it't}, \forall i \in \mathcal{I}, t \geq t' \in \mathcal{T}, \tag{9}$$

$$u_{it'(t-1)} \leq u_{it't}, \forall i \in \mathcal{I}, t \geq t'+1 \in \mathcal{T}, \tag{10}$$

$$c_{(i+1)t}(1 - x_{itt}) = 0, \forall i \in \mathcal{I}\backslash\{I\}, t \in \mathcal{T}, \tag{11}$$

$$u_{it'T} = 1, \forall i \in \mathcal{I}, t' \in \mathcal{T}, \tag{12}$$

$$x_{it't} \in \mathbb{B}, \forall i \in \mathcal{I}, t' \in \mathcal{T}, t \geq t' \in \mathcal{T}. \tag{13}$$

2.3.4 Objective function

The overall system cost is comprised of two cost components, i.e., MV operating cost and passenger waiting cost. The MV operating cost is the summation of the costs of all MV dispatches, i.e., $\sum_{l\in\mathcal{L},t\in\mathcal{T}} e_l y_{lt}$, where e_l is the operating cost for dispatching a level-l MV. The passenger waiting time at each time interval is calculated by the number of waiting passengers at each time interval (i.e. $p_{ijt'}(1-u_{it't})$) multiple the duration of each time interval (δ). Multiplying the time value for each passenger (i.e. w) and summing the passenger waiting time at each time interval together, we obtain the passenger waiting cost during the study time horizon as $\sum_{i\in\mathcal{I},j\in\mathcal{I}_i^+,t'\in\mathcal{T},t\geq t'\in\mathcal{T}} w\left[\delta p_{ijt'}(1-u_{it't})\right]$.

Then the objective function can be written as the summation of the passenger waiting cost and MV operating cost as shown in Equation (14).

$$\min_{y_{lt}, c_{it}, x_{it't}, u_{it't}} \sum_{l \in \mathcal{L}, t \in \mathcal{T}} e(l) y_{lt} + \sum_{i \in \mathcal{I}, j \in \mathcal{I}_i^+, t' \in \mathcal{T}, t \geq t' \in \mathcal{T}} w\left[\delta p_{ijt'} (1 - u_{it't})\right]. \tag{14}$$

2.3.5 Linearization

It can be seen except constraint (11), all the other constraints and the objective function are linear in the original formulation. The left-hand in constraint (11) is a bi-linear term involving the multiplication of a continuous variable and a binary variable. To simplify the model formulation to suit existing commercial solvers for integer linear programs (e.g. Gurobi), we equivalently replace constraint (11) with the following linear constraints (15). According to constraints (15), $x_{itt} = 0$ together with constraints (6) indicates that $c_{(i+1)t} = 0$, and $x_{itt} = 1$ indicates that $c_{(i+1)t}$ is constrained no more than constraints (6), which is exactly equivalent to constraints (11).

$$c_{(i+1)t} - LC x_{itt} \leq 0, \forall i \in \mathcal{I} \backslash \{I\}, t \in \mathcal{T}. \tag{15}$$

With that, the final model formulation is obtained as follows. It can be seen that the proposed MODOC problem is formulated with integer and continuous variables, and all the constraints and objective function are linear. Thus, the proposed problem is a MILP problem.

$$\begin{aligned} &\min_{y_{lt}, c_{it}, x_{it't}, u_{it't}} \sum_{l \in \mathcal{L}, t \in \mathcal{T}} e(l) y_{lt} + \sum_{i \in \mathcal{I}, j \in \mathcal{I}_i^+, t' \in \mathcal{T}, t \geq t' \in \mathcal{T}} w\left[\delta p_{ijt'} (1 - u_{it't})\right], \\ &\text{s.t.} \quad \text{Constraints } (1)-(10), (12), (13), (15). \end{aligned} \tag{16}$$

3. Solution approaches

The operation design problem in UTS is very computationally intensive, which may not be suitable real-world applications that demand a reasonable solution time. To overcome this challenge, this section aims to developing a solution algorithm that can efficiently obtain the optimal schedule for the investigated MODOC problem in an acceptable time. Section 3.1 identifies the theoretical upper and lower bounds to a feasible solution. Based on these two bounds and the clearance constraint (12), Section 3.2 proposes a customized DP algorithm to solve the investigated problem efficiently.

3.1 Theoretical properties

3.1.1 Upper bound to an optimal dispatch solution

Theorem 1. In an optimal solution to MODOC with dispatches $\{y_{lt}\}$ and vacant seats $\{c_{it}\}$, for any $t'' \in \mathcal{T}$, $l' \in \mathcal{L}$ with $y_{l't''} = 1$, then $c_{it''} < C, \exists i \in \mathcal{I}$.

Proof. This theorem is proven by contradiction. If the above statement does not hold, then $\exists t'' \in \mathcal{T}, l' \in \mathcal{L}$ with $y_{l't''} = 1$ and $c_{it''} \geq C, \forall i \in \mathcal{I}$. With this, we construct a new feasible solution with dispatches $\{\hat{y}_{lt}\}$, vacant seats $\{\hat{c}_{it}\}$, cumulative boarding percentages $\{\hat{u}_{it't}\}$ and boarding states $\{\hat{x}_{it't}\}$, where $\{\hat{y}_{lt}\}$ equals $\{y_{lt}\}$ except that $\hat{y}_{(l'-1)t''} = 1$ and $\hat{y}_{l't''} = 0$, and $\{\hat{c}_{it}\}$, $\{\hat{u}_{it't}\}$, $\{\hat{x}_{it't}\}$ can be uniquely solved by plugging $\{\hat{y}_{lt}\}$ into constraints (4)-(10),(12),(13),(15). This means that the new solution removes one MU from the MV dispatched at time t''. Because $c_{it''} \geq C, \forall i \in \mathcal{I}$, it indicates that in the optimal solution, all passenger groups at each station fully board the MV dispatched at time t'', and there are at least C seats without occupancy at each station. For example, if in an optimal dispatch solution, an MV with 4 MUs is dispatched. However, there is always C seats without occupancy at each station. Then an MV with 3 MUs also can satisfy the same passenger demand. After removing one MU from the MV dispatched at time t'' in the new solution, the vacant seats shall satisfy $\hat{c}_{it''} = c_{it''} - C, \forall i \in \mathcal{I}$. Further, it is easy to verify with constraints (4),(7)-(10),(12),(13),(15) that $\{\hat{u}_{it't}\} = \{u_{it't}\}$ and $\{\hat{x}_{it't}\} = \{x_{it't}\}$. Note that the new solution meets the passenger demand with the same passenger waiting cost. While the new solution incurs less MV operating cost than the original solution that dispatches one more MU. Thus, the new feasible solution even yields less overall system cost than the original solution, which contradicts the optimality of the original solution. This completes the proof. □

With Theorem 1, to guarantee the optimality of one solution, the number of MUs in the MV dispatched at the end of time interval $t \in \mathcal{T}$ must be less than the minimum $l' \coloneqq \sum_{l \in \mathcal{L}} l \cdot y_{lt}$ that has $c_{it} \geq C, \forall i \in \mathcal{I}$. Thus, the set of values of l' at time intervals $t \in \mathcal{T}$ serves as an upper bound to the optimal dispatch solution.

3.1.2 Lower bound to an optimal dispatch solution

Theorem 2. With an optimal solution to MODOC with dispatches $\{y_{lt}\}$ and vacant seats $\{c_{it}\}$, if $\exists t'' \in \mathcal{T}, l' \in \mathcal{L}$ with $y_{l't''} = 1$, and $c_{it''} = 0, \forall i \in \mathcal{I} \backslash \{1\}$. Then we construct a new feasible solution with dispatches $\{\hat{y}_{lt}\}$ obtained by copying $\{y_{lt}\}$ except for setting $\hat{y}_{l''t''} = 1$ and $\hat{y}_{l't''} = 0$ for any $l'' \in \{l+1, l+2, \cdots, L\}$ (which means to only increase the MV capacity level dispatched at time t'' while keeping the rest the same), and vacant seats $\{\hat{c}_{it}\}$, cumulative boarding percentage $\{\hat{u}_{it't}\}$, and boarding state $\{\hat{x}_{it't}\}$ obtained from constraints (4)-(10),(12),(13),(15), the new solution must satisfy $\hat{c}_{it''} > 0, \exists i \in \mathcal{I} \backslash \{1\}$.

Proof. This theorem is also proven by contradiction. If the above statement does not hold, then $\hat{c}_{it''} = 0, \forall i \in \mathcal{I} \backslash \{1\}$ in the new feasible solution.

To visualize the conversion from the original optimal solution to the new solution, we transform $\{y_{lt}\}$ to a $1 \times T$ array such that $\{Y_t\}$ and $Y_t = \sum_{l \in \mathcal{L}} l y_{lt} \;\; \forall t \in \mathcal{T}$. Now $\{Y_t\}$ indicates whether an MV is dispatched or not at any time t (i.e., yes if $Y_t > 0$ or no if $Y_t = 0$) and how many MUs are dispatched at time t (i.e., the value of Y_t). Thus, $\{Y_t\}$ and $\{y_{lt}\}$ are equivalent and transferable from one to another. Further, we transform $\{\hat{y}_{lt}\}$ to $\{\hat{Y}_t\}$ in the same way. As illustrated in the toy example in Fig. 5, in the original optimal solution, a level-3 MV is dispatched at time $t'' = 5$ (i.e., $Y_5 = l' = 3$), while in the new solution, a longer level-6 MV is dispatched instead (i.e., $\hat{Y}_5 = l'' = 6$) yet the rest remains the same. Note that with the above assumption, we know that $c_{it''} = \hat{c}_{it''} = 0, \forall i \in \mathcal{I} \backslash \{1\}$.

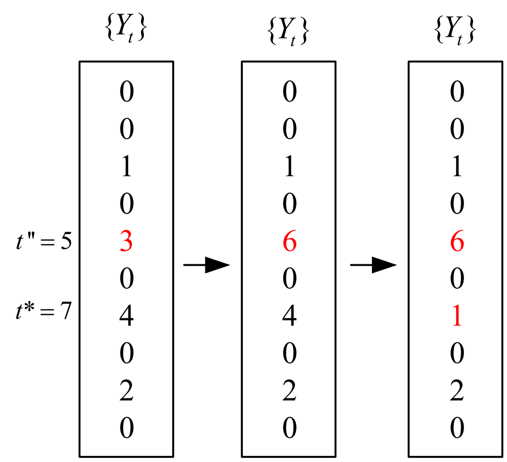


Fig. 5 Solution construction process.

With this, we construct another solution with dispatches $\{\tilde{y}_{lt}\}$ by integrating $\{y_{lt}\}$ and $\{\hat{y}_{lt}\}$. Again, we use $\{\tilde{Y}_t\}$ to equivalently represent $\{\tilde{y}_{lt}\}$ after the same transformation. Since $c_{it''} = \hat{c}_{it''} = 0, \forall i \in \mathcal{I}\backslash\{1\}$, then we know that in solution $\{y_{lt}\}$, at minimum $(l'' - l')C$ passengers are waiting to board at each station after time t'', which will be cleared by later dispatches in $\{y_{lt}\}$ to meet clearance constraint (12). This indicates that $\exists t^* > t''$ such that $\sum_{t=t''+1}^{t^*} Y_t \geq l'' - l'$ and $\sum_{t=t''+1}^{t^*-1} Y_t < l'' - l'$. Then we can construct $\{\tilde{y}_{lt}\}$ by setting $\tilde{Y}_{t''} = \hat{Y}_{t''}$ and $\tilde{Y}_t = 0, \forall t'' < t < t^*$, $\tilde{Y}_{t^*} = \sum_{t=t''+1}^{t^*} Y_t - (l'' - l')$ and $\tilde{Y}_t = Y_{t''}, \forall t < t''$or $> t^*$. As illustrated in Fig. 5, in the solution $\{\tilde{Y}_t\}$, a level-6 MV is dispatched at time $t'' = 5$ (i.e., $\tilde{Y}_5 = 6$) and a level-1 MV is dispatched at time $t^* = 7$ (i.e., $\tilde{Y}_7 = 1$) instead of dispatching a level-4 MV at that time. The corresponding $\{\tilde{c}_{it}\}$, $\{\tilde{u}_{it't}\}$, and $\{\tilde{x}_{it't}\}$ values are calculated by plugging $\{\tilde{y}_{lt}\}$ into equation (17)-(19).

$$\begin{cases} \tilde{c}_{1t} = \sum_{l \in \mathcal{L}} lC\tilde{y}_{lt}, \forall t \in \mathcal{T}, \\ \tilde{c}_{it} = \sum_{l \in \mathcal{L}} lC\tilde{y}_{lt} + \sum_{i'=1}^{i-1} \left(\sum_{j \in \mathcal{I}_{i'}^{-}, t' \leq t \in \mathcal{T}} \left(\tilde{u}_{jt't} - \tilde{u}_{jt'(t-1)}\right) p_{ji't'} - \sum_{j \in \mathcal{I}_{i'}^{+}, t' \leq t \in \mathcal{T}} \left(\tilde{u}_{i't't} - \tilde{u}_{i't'(t-1)}\right) p_i \right. \\ \forall i \in \mathcal{I}\backslash\{1\}, t \in \mathcal{T}\backslash\{1\}, \end{cases} \tag{17}$$

$$\tilde{u}_{it't} = \begin{cases} \tilde{u}_{it'(t-1)}, if \sum_{l \in \mathcal{L}} \tilde{y}_{lt} = 0; \\ 1, if\ A \leq B; \\ \begin{cases} 1, \forall t' < t \\ E, \forall t' = t' \end{cases} if\ A > B\ and\ D \leq E; \\ \begin{cases} 1, if\ t' < t^s \\ 0, if\ t' > t^s \\ F, if\ t' = t^s \end{cases}, if\ A > B\ and\ D > E, \end{cases} \tag{18}$$

where

$$A := \sum_{j\in\mathcal{J}_i^+, t'\le t-1\in\mathcal{T}} \Big(\big(1-\tilde{u}_{it'(t-1)}\big)p_{ijt'} + p_{ijt}\Big),$$
$$B := \tilde{c}_{it} + \sum_{j\in\mathcal{J}_i^-, t'\le t\in\mathcal{T}} \big(u_{jt't} - u_{jt'(t-1)}\big)p_{jit'},$$
$$D := \sum_{j\in\mathcal{J}_i^+, t'\le t-1\in\mathcal{T}} \big(1-\tilde{u}_{it'(t-1)}\big)p_{ijt'},$$
$$E := \left(\sum_{t'=1}^{t-1}\big(1-u_{it'(t-1)}\big)p_{ijt'} + p_{ijt} - \tilde{c}_{it} - \sum_{j\in\mathcal{J}_i^-, t'\le t\in\mathcal{T}} \big(u_{jt't} - u_{jt'(t-1)}\big)p_{jit'}\right)/p_{ijt},$$
$$F := \left(\sum_{t'=1}^{t^s}\big(1-u_{it'(t-1)}\big)p_{ijt'} - \tilde{c}_{it} - \sum_{j\in\mathcal{J}_i^-, t'\le t\in\mathcal{T}} \big(u_{jt't} - u_{jt'(t-1)}\big)p_{jit'}\right)/p_{ijt^s},$$

and t^s is the arriving time of passenger group that cannot fully board the MV dispatched at time t. t^s can be obtained by $\sum_{t'=1}^{t^s-1}\big(1-u_{it'(t-1)}\big)p_{ijt'} \le \tilde{c}_{it} + \sum_{j\in\mathcal{J}_i^-, t'\le t\in\mathcal{T}}\big(u_{jt't} - u_{jt'(t-1)}\big)p_{jit'}$ and $\sum_{t'=1}^{t^s}\big(1-u_{it'(t-1)}\big)p_{ijt'} > \tilde{c}_{it} + \sum_{j\in\mathcal{J}_i^-, t'\le t\in\mathcal{T}}\big(u_{jt't} - u_{jt'(t-1)}\big)p_{jit'}$.

$$\tilde{x}_{it't} = \begin{cases} 0, if\ \tilde{u}_{it't} \ne 1 \\ 1, if\ \tilde{u}_{it't} = 1 \end{cases}, \forall i \in \mathcal{I}, t \ge t' \in \mathcal{T}. \tag{19}$$

Because $\tilde{c}_{it''} = 0, \forall i \in \mathcal{I}\backslash\{1\}$ and $\tilde{Y}_{t''} + \tilde{Y}_{t^*} = Y_{t''} + Y_{t^*}$, it indicates we dispatch $l'' - l'$ MUs $t^* - t''$ time intervals earlier than the optimal solution. By plugging $\{\tilde{c}_{it}\}$, $\{\tilde{u}_{it't}\}$, and $\{\tilde{x}_{it't}\}$ of the constructed solution to constraints (4)-(10),(12),(13),(15), all constraints are satisfied. Thus, this newly constructed solution is feasible. Then, it is obvious that more passenger waiting cost will be occurred by the optimal solution compared with the newly constructed solution. Further, we mentioned that the operating cost function $e(l)$ is concave over the level of MV (i.e., l), and the operating cost of dispatching a vehicle with C seats is far less than the passenger waiting cost if we let the corresponding number of passengers wait for one more time interval (i.e., $e(1) \ll w\delta C$). Thus the overall system cost of dispatching $\tilde{Y}_{t''}$ and $\tilde{Y}_{t^*}$ MUs at time t'' and t^* is less than that of dispatching $Y_{t''}$ and Y_{t^*} MUs at time t'' and t^*. The optimal solution has higher overall system cost than the constructed solution, which contradicts the optimality of the original solution. This completes the proof. □

With Theorem 2, to guarantee the optimality of one solution, the number of MUs in the MV dispatched at the end of time interval $t \in \mathcal{T}$ must be greater than or equal to the maximum $l' \coloneqq \sum_{l\in\mathcal{L}} l \cdot y_{lt}$ that has $c_{it} = 0, \forall i \in \mathcal{I}\backslash\{1\}$. Thus, the set of values of l' at time intervals $t \in \mathcal{T}$ serves as a lower bound to the optimal dispatch solution.

3.2 Dynamic Programming Algorithm

This section proposes a customized DP algorithm to solve the investigated MODOC problem. The outstanding challenge to the investigated problem is that the passenger queues at each time interval are originally unbounded, which much increases the state space and thus increases the problem complexity.

Thanks to the theoretical properties studied in Section 3.1, which provides relatively tight upper and lower bounds to a feasible solution, the state space can be largely reduced to a narrow band at each stage. Tab. 2 lists the basic elements of the customized DP algorithm, and the explicit explanation of each element follows.

Tab. 2 Basic elements of the customized DP algorithm.

No.	Elements	Notation	No.	Elements	Notation
1	Stage	$t \in \{0\} \cup \mathcal{T}$	4	MU terminal stock state	o_t
2	Decision variable	Y_t	5	Boarding position state	v_t
3	MU pipeline stock state	g_t	6	Cost function	M_t

1) Stages

Consider each discrete time interval $t \in \mathcal{T}$ as a stage of the studied problem. Set $t = 0$ is a dummy stage for the convenience of the notations.

2) Decision variables

The numbers of dispatched MUs at Stage t, denoted as Y_t, is the decision variable in the customized DP algorithm. It is a nonnegative integer no greater than the MU terminal stock (i.e., o_t; see the next section for the definition) and the maximum number of MUs that a MV can have (i.e., L), i.e.,

$$Y_t \in \{0,1,\cdots,\min(o_t, L)\}, \forall t \in \mathcal{T}. \tag{20}$$

3) States

There are three sets of state variables in the proposed DP algorithm, and they are the MU pipeline stock, the MU terminal stock, and the boarding positions of the passenger groups.

The MU pipeline stock at a given Stage t, denoted by $g_t \coloneqq [g_{tt'}]_{\forall t' \in \{1,2,\cdots,P\}}$, is an $1 \times P$ array with non-negative integer elements. $g_{tt'}$ denotes the number of MUs that will return to Station 1 at the beginning of Stage $t + t'$. For example, if g_{t1} equals 0, no MU returns to Station 1 at the beginning of Stage $t + 1$; otherwise, the value of g_{t1} means the number of MUs returning to Station 1 at the beginning of Stage $t + 1$. The initial values of $[g_{tt'}]$ are set to 0 at Stage 0, since all vehicles are assumed to be stored at Station 1 before operations and thus no vehicles are being circulated across other stations to return to Station 1 at the beginning.

The MU terminal stock state, denoted by o_t, pertains to the maximum number of MUs that can be dispatched from Station 1 at the beginning of Stage t. It is a nonnegative integer, and the initial value of o_t is set to V at Stage 0.

The boarding positions of the passenger groups, denoted by $v_t \coloneqq [v_{it}]_{\forall i \in \mathcal{I}}$, mark the last arrival time of boarded passengers at each Station i at the end of Stage t. Note that due to the MSFIFO, at each Stage t, all passenger groups with different destinations at the same origin Station i have the same boarding position, and thus we do not need to track queues with different destinations separately, which much reduces the state space. v_{it} is a nonnegative real number. The value of $\lfloor v_{it} \rfloor$ denotes the arrival time interval of the last group of passengers who has boarded at Station i at the end of Stage t. The value of $v_{it} - \lfloor v_{it} \rfloor$ indicates the boarding portion of the passenger group arriving at time interval $\lfloor v_{it} \rfloor + 1$ at the end of Stage t. For example, $v_{it} = 7.3$ indicates that the passenger groups arriving at Station i at time interval $\{1,2,\cdots,7\}$ have all boarded MVs at the end of Stage t. The boarding portion of each passenger group arriving at Station i at time interval 8 is 0.3 (or 30%) at the end of Stage t. The initial values of $\lfloor v_{it} \rfloor$ are set to 0 at stage 0.

4) State transition function

The state space at each Stage t is denoted by $S_t \coloneqq \{(g_t, o_t, v_t)\}$. The state transition function is shown in Equation (21). By plugging the states (i.e., $S_{t-1} \coloneqq \{(g_{t-1}, o_{t-1}, v_{t-1})\}$) and decision variable (i.e., Y_{t-1}) at Stage $t-1$ into Equation (21), the states at Stage t (i.e., $S_t \coloneqq \{(g_t, o_t, v_t)\}$) can be obtained.

$$
\begin{aligned}
&T\left(\begin{bmatrix} g_{t-1} \\ o_{t-1} \\ v_{t-1} \end{bmatrix}, Y_{t-1}\right) = \begin{bmatrix} g_t \coloneqq \left[g_{tt'} = g_{(t-1)(t'+1)}, \forall t' \in \{1,2,\cdots,P-1\}, g_{tP} = Y_{t-1}\right] \\ o_t \coloneqq o_{t-1} + g_{(t-1)1} - Y_{t-1} \\ v_t \coloneqq \begin{bmatrix} v_{1t} = arg \max_{t'' \in [\lfloor v_{1(t-1)} \rfloor, t]} G(1,t'') \geq c_{1t} \coloneqq Y_{t-1}C \\ v_{it} = arg \max_{t'' \in [\lfloor v_{i(t-1)} \rfloor, t]} G(i,t'') \geq c_{it} + a(i,t), \forall i \in \mathcal{I} \backslash \{1\} \end{bmatrix} \end{bmatrix}, \forall t \\
&\in \mathcal{T},
\end{aligned}
$$

where

$$
\begin{aligned}
&c_{it} := \max\left\{c_{(i-1)t} + a(i-1,t) - G(i-1,t), 0\right\}, \forall i \in \mathcal{I} \backslash \{1\}, t \in \mathcal{T}, \\
&a(i,t) \coloneqq \sum_{i' \in \mathcal{I}_i^-} \left[\sum_{t' = \lfloor v_{i'(t-1)} \rfloor + 1}^{\lfloor v_{i't} \rfloor} \min\{(v_{i't} - t'), 1\} p_{i'it'} \right. \\
&\qquad \left. + \left(\lfloor v_{i'(t-1)} \rfloor + 1 - v_{i'(t-1)}\right) p_{i'i\lfloor v_{i'(t-1)} \rfloor + 1} \right], \forall i \in \mathcal{I} \backslash \{1\}, t \in \mathcal{T}, \\
&a(1,t) = 0, \forall t \in \mathcal{T}, \\
&G(i', t'') := \sum_{j \in \mathcal{I}_{i'}^+} \left[(t'' - \lfloor t'' \rfloor) p_{i'j(\lfloor t'' \rfloor + 1)} + \sum_{t' = \lfloor v_{i(t-1)} \rfloor + 1}^{\lfloor t'' \rfloor} \min\{(t' - v_{i(t-1)}), 1\} p_{i'jt'} \right], \forall i' \\
&\in \mathcal{I}, t'' \in [\lfloor v_{i(t-1)} \rfloor, t].
\end{aligned}
\tag{21}
$$

Consistent with Section 2, c_{it} denotes the amount of vacant seats of the MV dispatched at the end of Stage t on arriving at Station i. Function $a(i,t)$ denotes the number of alighting passengers when the MV dispatched at the end of Stage t arrives at Station i. The number of alighting passengers at Station 1 is set to 0 for each MV (i.e., $a(1,t) = 0, \forall t \in \mathcal{T}$). Function $G(i', t'')$ denotes the number of the passengers waiting at Station i' at Time t''. We slightly abuse the notation to allow t'' to be a fractional number. This allows us to track passenger queues over the continuous time. Note that to generalize this function, the range of Time t'' is $[\lfloor v_{i(t-1)} \rfloor, t]$ that indicates the time index can be real number. The detailed calculation process of state v_t can refer to Equations (18).

5) Customized DP algorithm

With these definitions, the customized DP algorithm is performed as the following steps, and the pseudo-code is shown in Algorithm.

Step 1: Start from Stage $t = 0$ with state space $S_0 \coloneqq \{g_0, o_0, v_0\}$ and cost $M_0 \coloneqq 0$. Set $t = 1$.

Step 2: Set state space $S_t = \emptyset$.

Step 2.1: For each state $s_{t-1} \coloneqq (g_{t-1}, o_{t-1}, v_{t-1}) \in S_{t-1}$, state s_t is obtained by the transition function (i.e. $s_t = T(s_{t-1}, Y_{t-1})$). If any of the following conditions (i.e., upper and lower bounds and clearance constraint (12)) is met for this state, go to next state in S_{t-1} and repeat Step 2.1; otherwise, go to next step.

$$
\begin{gathered}
Y_t \neq 0, c_{it} \geq C, \forall i \in \mathcal{I}; \\
\text{or } Y_t \neq 0, c_{it} = 0, \forall i \in \mathcal{I} \backslash \{1\} \text{ and } \hat{c}_{it} = 0, \forall i \in \mathcal{I} \backslash \{1\}; \\
\text{or } v_{iT} \neq T, \forall i \in \mathcal{I}.
\end{gathered}
\tag{22}
$$

where the definition of $\hat{c}_{it}$ is consistent with Section 3.1.2.

Step 2.2: Calculate the cost of $M'_t(s_t \coloneqq T(s_{t-1}, Y_{t-1})) := M^*_{t-1}(s_{t-1}) + e(Y_{t-1}) + \sum_{i \in \mathcal{I}, j \in \mathcal{I}_i^+} w(\sum_{t'=\lfloor v_{it} \rfloor + 1}^{t} \min\{(t' - v_{it}), 1\} p_{ijt'})$, where $e(Y_{t-1})$ is the cost for dispatching Y_{t-1} MUs. If $s_t \notin S_t$, add s_t to S_t, and set mappings $M^*_t(s_t) = M'_t(s_t)$, $Y^*_{t-1}(s_t) = Y_{t-1}$, and $S^*_{t-1}(s_t) = s_{t-1}$; otherwise, if $M'_t(s_t) < M^*_t(s_t)$, update mappings $M^*_t(s_t) = M'_t(s_t)$, $Y^*_{t-1}(s_t) = Y_{t-1}$, and $S^*_{t-1}(s_t) = s_{t-1}$.

Step 2.3: If $t < T$, $t = t + 1$, go to Step 2; otherwise, go to Step 3.

Step 3: Get mappings at Stage T with the minimum cost in the state space (i.e., $\min\{M^*_T(s_T)\}, \forall s_T \in S_T$), and set it as the optimal cost. Denote the state for this optimal cost by s^*_T. Then the optimal dispatch solution $\{Y^*_t\}$ is obtained by tracking back from Stage T to Stage 1 according to $Y^*_{t-1} = Y^*_{t-1}(s^*_t)$ and $s^*_{t-1} = S^*_{t-1}(s^*_t)$.

Algorithm. Dynamic Programming Algorithm

Input $\mathcal{I}; \mathcal{T}; \mathcal{L}; \delta; C; V; P; H; e_l, \forall l \in \mathcal{L}; p_{ijt'}, \forall i \in \mathcal{I}, j \in \mathcal{I}_i^+, t' \in \mathcal{T}$
$t \leftarrow 0, S_0 \leftarrow \{g_0, o_0, v_0\}, M_0 \leftarrow 0$
For $t = 1$ to $\mathcal{T}$ **do**
  $S_t \leftarrow \emptyset$
  For each state $s_{t-1} \coloneqq (g_{t-1}, o_{t-1}, v_{t-1}) \in S_{t-1}$ **do**
    $s_t = T(s_{t-1}, Y_{t-1})$

If s_t does not meet any of the conditions in Equation (22)
Calculate the cost of $M'_t\big(s_t \coloneqq T(s_{t-1}, Y_{t-1})\big)$
If $s_t \notin S_t$
add s_t to S_t, set mappings $M^*_t(s_t) = M'_t(s_t)$, $Y^*_{t-1}(s_t) = Y_{t-1}$, and $S^*_{t-1}(s_t) = s_{t-1}$
Else if $M'_t(s_t) < M^*_t(s_t)$,
update mappings $M^*_t(s_t) = M'_t(s_t)$, $Y^*_{t-1}(s_t) = Y_{t-1}$, and $S^*_{t-1}(s_t) = s_{t-1}$
End if
End if
End for
End for
Get $\min\{M^*_T(s_T)\}, \forall s_T \in S_T$. Denote the state for this optimal cost by s^*_T
For $t = T$ to 1 **do**
$Y^*_{t-1} = Y^*_{t-1}(s^*_t)$, $s^*_{t-1} = S^*_{t-1}(s^*_t)$
End for
Output The optimal dispatch solution $\{Y^*_t\}$, the optimal objective $\min\{M^*_T(s_T)\}, \forall s_T \in S_T$

Note that due to the proposed upper and lower bounds in Step 2.1, the number of states in each stage can be largely reduced. If we drop the upper and lower bounds in Step 2.1 from the proposed customized DP algorithm, the algorithm will become the original DP algorithm, and a greater number of states will remain in each stage, which significantly increase the computational time. The detailed comparisons between the customized and original DP algorithms can be found in the numerical experiments.

4. Numerical experiments

In this section, several numerical experiments are conducted to test the performance of the proposed model and algorithm, including a case study based on a hypothetical example and a real-world case study based on the passenger data obtained from Beijing Subway Line 6. The proposed customized DP algorithm is coded in Visual studio C++ 2017 at a Window 7 PC with i7-4790 GPU and 16.0 GB RAM. The benchmark instances for comparison are solved by a commercial solver, Gurobi 650, at the same platform.

4.1 Case study on a hypothetical example

In the hypothetical example, at default, we consider a single direction corridor with 4 stations. MVs start at Station 1 and destine for Station 4 (i.e., I=4). The number of MUs in stock is 50 (i.e., V=50) and each MU can accommodate 20 persons (i.e., C=20). The maximum number of MUs for one MV is limited as 6 (i.e., L=6). In addition, the maximum study time horizon is 60 time intervals (i.e., T=60), and the passenger demand in the skewed time coordinates is shown in Tab. 3. Coefficient w is set as 0.11\$/minute, and level-l MV operating cost e_l is $2.049\$ + 0.37\$ * (l * C)^{0.5}$ in consistence with Chen et al., (2019a). Without losing generality, three different demand scenarios are taken into consideration, including the off-peak hour scenario (1-10 time intervals, 41-55 time intervals), the peak hour scenario (11-20 time intervals, 31-40 time intervals) and the transition hour scenario (21-30 time intervals).

Tab. 3 The arriving passenger demand for study time horizon.

Time interval	Arriving passenger demand
1-10	10 passengers/ time interval
11-20	40 passengers/ time interval

21-30	30 passengers/ time interval
31-40	40 passengers/ time interval
41-55	10 passengers/ time interval
56-60	0 passengers/ time interval

With the default settings above, we derive a set of instances based on the different combinations of the time horizon, headway and operational cycle. The instance index includes the time horizon (T), headway (H) and operational cycle (P) for this specific instance, i.e., IN-T-H-P. For example, index IN-20-1-5 indicates that the study time horizon of this instance is 20 time intervals, and the headway and operational cycle are 1 time interval and 5 time intervals, respectively.

We compare the solutions from Gurobi and the customized DP on objective value, computation time, gap, number of dispatched MVs and number of dispatched MUs. The maximum CPU time is limited to 300s. Tab. 4 shows the comparison results for a total of 12 instances. It is found that as the scale of the instance increases, the Gurobi solver fails to obtain the optimal solution. However, the customized DP algorithm successfully solve all the instances within the time limit. The customized DP algorithm has a solution time much less than that from Gurobi in each instance, which indicates the superior solution efficiency of the proposed customized DP algorithm. Especially, in some instances such as IN20-2-5 and IN-40-2-5, the proposed customized DP algorithm is 200-500 times faster than Gurobi. Note that due to the considerations of returns of MV/MUs, the numbers of the dispatched MUs may be greater than the total number of MUs in stock (i.e., 50) if some MUs are returned and dispatched multiple times. However, the total number of dispatched MUs in each operational cycle (i.e., P) is always restricted by Constraints (1).

Tab. 4 The comparison for different instances with Gurobi, the original DP and customized DP.

Instance parameter	Method	Objective ($)	Computation time	Gap	# of MVs	# of MUs
IN-20-1-5	Gurobi Customized DP	151.4	23.05s 0.06s	0%	10	25
IN-20-1-10	Gurobi Customized DP	151.4	22.00s 5.86s	0%	10	25
IN-20-2-5	Gurobi Customized DP	172.8	13.37s 0.02s	0%	7	26
IN-20-2-10	Gurobi Customized DP	172.8	12.90s 0.55s	0%	7	26
IN-40-1-5	Gurobi Customized DP	344.3	107.20s 0.78s	0%	20	60
IN-40-1-10	Gurobi Customized DP	344.3	106.62s 80.51s	0%	20	60
IN-40-2-5	Gurobi Customized DP	415.2	58.98s 0.22s	0%	13	61
IN-40-2-10	Gurobi Customized DP	415.2	59.21s 5.28s	0%	13	61
IN-60-1-5	Gurobi	482	300.00s	6.29%	26	85

	Customized DP	478.3	3.69s	0%	27	84
IN-60-1-10	Gurobi	482	300.00s	6.29%	26	85
	Customized DP	478.3	290.10s	0%	27	84
IN-60-2-5	Gurobi	580.6	300.00s	4.93%	18	88
	Customized DP	561.3	0.70s	0%	19	96
IN-60-2-10	Gurobi	580.6	300.00s	4.93%	18	88
	Customized DP	561.3	19.31s	0%	19	96

Based on the obtained results, it is known that the proposed customized DP algorithm performs much better than the Gurobi solver on the solution time. It benefits from the proposed upper and lower bounds. Fig. 6 compares the state space between the original DP algorithm and the customized DP algorithm based on IN-40-1-5. Note that the state space of the original DP algorithm rapidly increases due to the unbounded passenger queue at each stage. However, for the customized DP algorithm, the state space curve has a much lower increase rate because the tight bounds and are applied at each stage, which dramatically reduces the state space and thus the solution time.

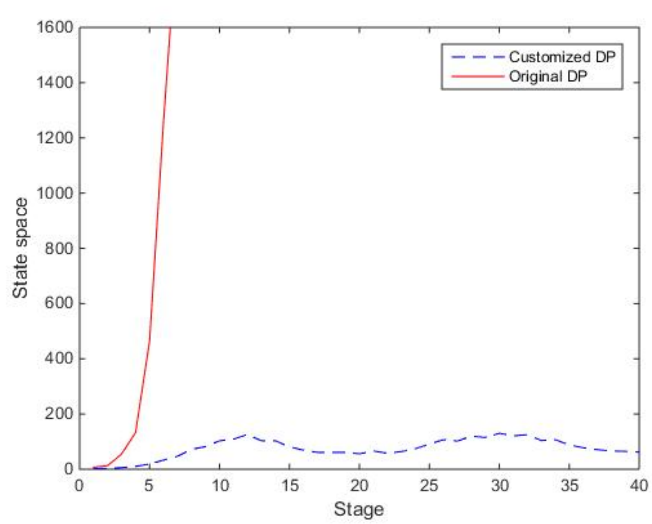


Fig. 6 Comparison of the state space between the original and customized DP algorithms.

4.2 Case Study on Beijing UTS Line 6

4.2.1 Experiment description and parameter settings

Beijing UTS Line 6 is a corridor line in Beijing, China. It is the first rail transit line in Beijing UTS with 8 B-type units, which accommodates as many as 1840 people. The total length of Line 6 is about 53 km consisting 3 phases, where Phase 1 is 30km, Phase 2 or eastern extension is 12km and Phase 3 or western extension is 11km. Since Phase 1 is the longest and carries the most traffic, the case study is conducted on Phase 1 as shown in Fig. 7.

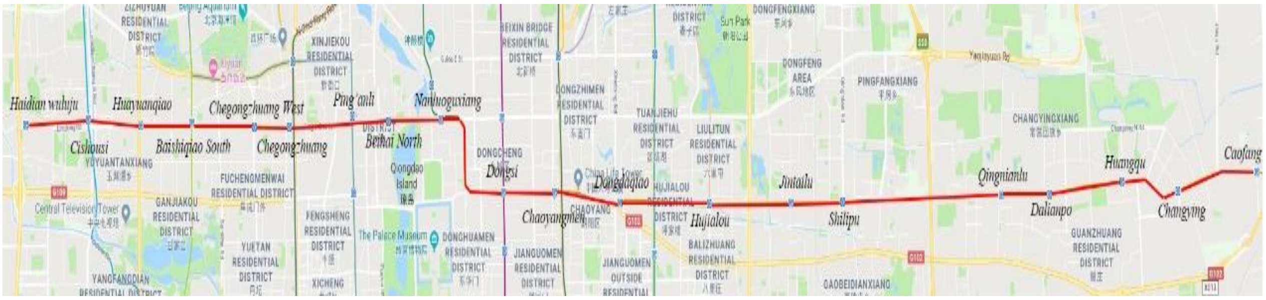


Fig. 7 Beijing UTS Line 6, Phase 1 (source: Google Maps).

In the daily operations of the investigated UTS, the planned operational cycle (P) and minimal headway (H) are 16 min and 4 min, respectively (https://www.wikipedia.org/). Phase 1 contains 20 stations starting from the Haidian Wuluju station and ending at the Caofang station. The passenger demand data (with a resolution of one minute) at 19 out of the 20 stations (except the Chegongzhuang station where the data is missing) on Nov 11, 2017 are used in the case study. The total passenger arrival rate across the 19 stations is shown in Fig. 8. Three time periods are selected as shown on the white areas in Fig. 8, including 7:40 am-8:40 am for the off-peak hour scenario, 11:00 am-12:00 pm for the transition hour scenario, and 19:00 pm-20:00 pm for the peak hour scenario.

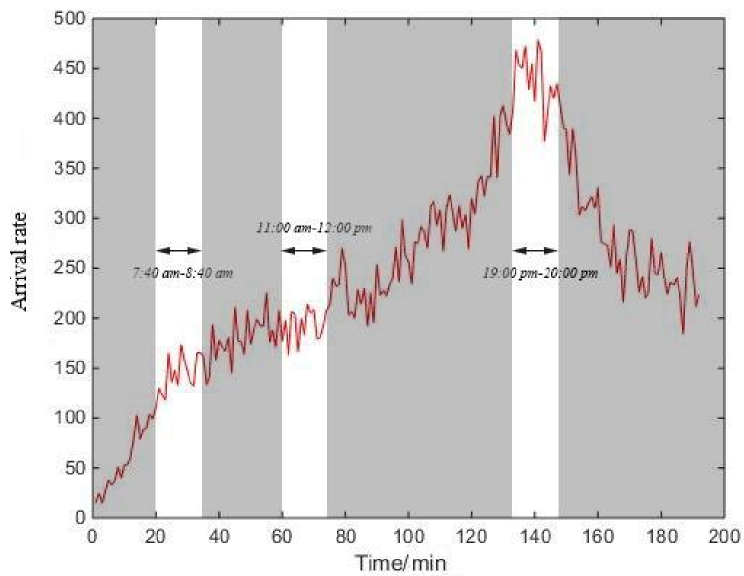


Fig. 8 Line 6 total arrival rate.

The default parameter values are listed in Tab. 5.

Tab. 5 Default parameter values for the case study.

Parameters	Symbol	Value
Number of stations	I	19
Number of time intervals	T	60
Length of one time interval	δ	1 minutes

Capacity of one MU	C	230 (B type unit)
MUs in stock	V	50
Minimum headway	H	4 minutes
Operational cycle	P	16 minutes
Energy and associated cost in dispatching l MUs	e_l	$e_l = 2.049\$ + 0.37\$ * (l * C)^{0.5}$
Value of waiting time for each passenger	w	0.11$/minute

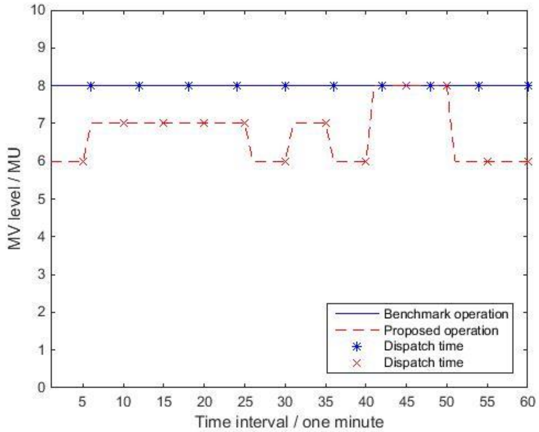
4.2.2 Computational results

Due to the increased instance size, both the original DP algorithm and the Gurobi solver cannot solve a feasible solution after several hours. Hence, this section only shows the results of the customized DP algorithm.

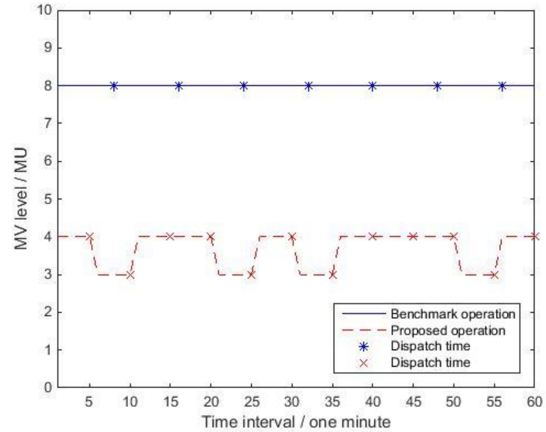
Tab. 6 Results of the customized DP algorithm.

Period type	Objective ($)	Computation time (s)	# of MVs	# of MUs
Off-peak hour 7:40 am-8:40 am	2585.3	79.6	12	34
Transition hour 11:00 am-12:00 pm	4217.1	235.3	12	44
Peak hour 19:00 pm-20:00 pm	7174.2	572.5	12	81

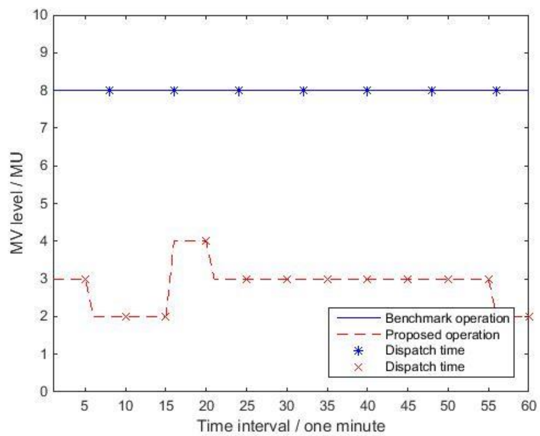
As shown in Tab. 6, all the instances are solved in acceptable computation times suitable for implementation. To compare the proposed operation with the existing operation, we construct a benchmark instance with fixed-capacity vehicles and fixed dispatch headways. Per the existing schedule (https://www.bjsubway.com/), in the benchmark instance, we set the dispatch headways for peak hours and off-peak hours as 5 minutes and 7 minutes, respectively, and we fix the vehicle level at 8. Fig. 9 shows the comparison results. In Fig. 9 (a)-(c), the blue curve with the star marks plots the levels and the times of the dispatched MVs in the corresponding scenario in the benchmark operation, and the red curve with the cross marks plots those in the proposed operation. We see that the proposed operation prefers to dispatch shorter MVs more frequently compared with the benchmark, particularly in the off-peak hour scenario. To measure the improvement of the proposed operation, Tab. 7 compares the overall system cost (OSC), vehicle operating cost (VOC), and average waiting time (AWT) between the benchmark and the proposed operations. Note that the proposed operation performs better than the benchmark operation for the different scenarios. Particularly, for the off-peak hour scenario, the proposed operation reduces the overall system cost and the average waiting time by 45.3% and 40.8%, respectively.



(a) Peak hour scenario



(b) Transition hour scenario



(c) Off-peak hour scenario

Fig. 9 Schedule comparison between the benchmark and the proposed operation.

Tab. 7 Objective comparison between the benchmark and the proposed operation.

Period type	Proposed operation			Benchmark operation		
	OSC ($)	VOC ($)	AWT (min)	OSC ($)	VOC ($)	AWT (min)
Off-peak hour	2585.3	137.4	2.00	4722.4	125.4	3.38
Transition hour	4217.1	153.2	2.00	7018.7	125.4	3.39
Peak hour	7174.2	199.3	2.23	9145.5	179.2	2.57

To show the effectiveness of the proposed MSFIFO rule, Fig. 10 plots the passenger queueing and boarding behaviors of the first five time intervals at the 8th station (i.e., Beihai North Station). Three types of passenger groups are described in the figure, including waiting passenger groups (solid square), boarded passenger groups (colored square), and oncoming passenger groups (dash square; the time index marks the time interval at which they will arrive).

In Fig. 10 (a), the following process is described: At time interval 1, passenger groups arrive at the station; one MV arrives; Only a portion (78%) of the passenger groups board the MV, and the remaining passengers need to wait for the next MV. In Fig. 10 (b)-(e), passenger groups keep arriving at the station yet cannot board a MV until time interval 5 when the next MV arrives. However, as shown in Fig. 10 (e), due to the limited MV capacity, not all passengers can board. Based on the proposed MSFIFO rule, the passenger groups arriving earlier have a higher priority to board the MV. Thus, only the passenger groups arriving at time interval 1 through 4 fully board the MV. However, only a portion (93%) of the passenger groups arriving at time interval 5 board the MV, and the remaining passengers need to wait for the next MV. Note that there are no queue jumpers in all these queues, and passengers arriving at the same time have the same boarding priority, which is consistent with the proposed MSFIFO rule.

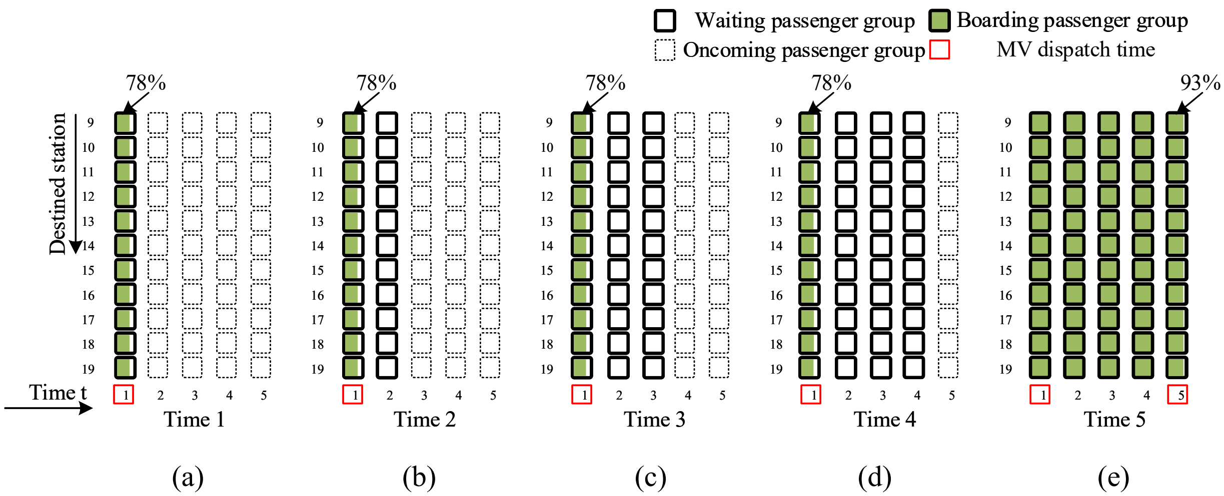
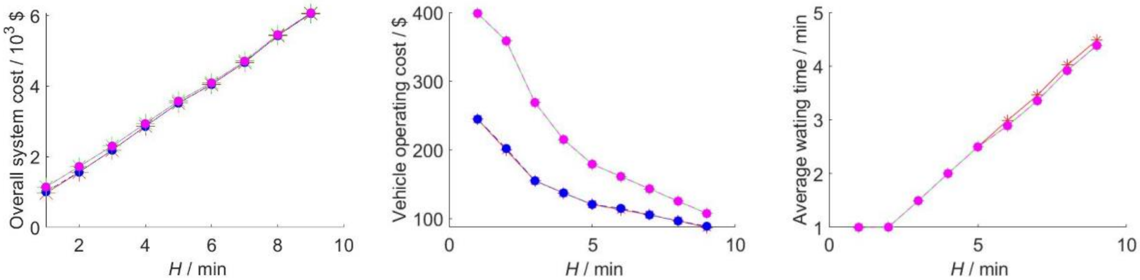


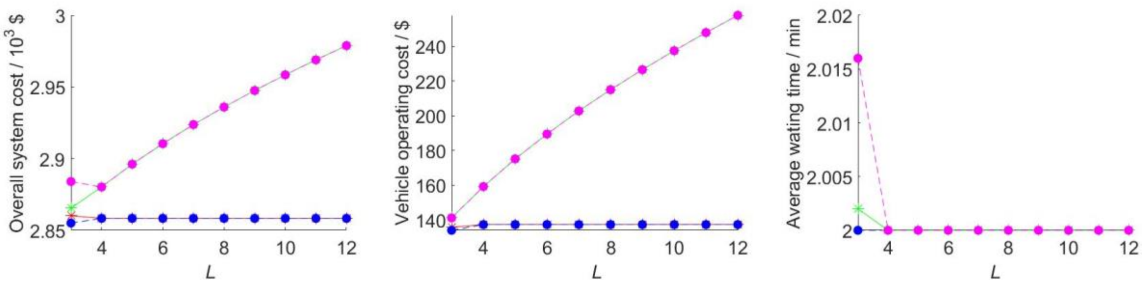
Fig. 10 Illustration of the MSFIFO rule.

4.2.3 *Sensitivity analysis*

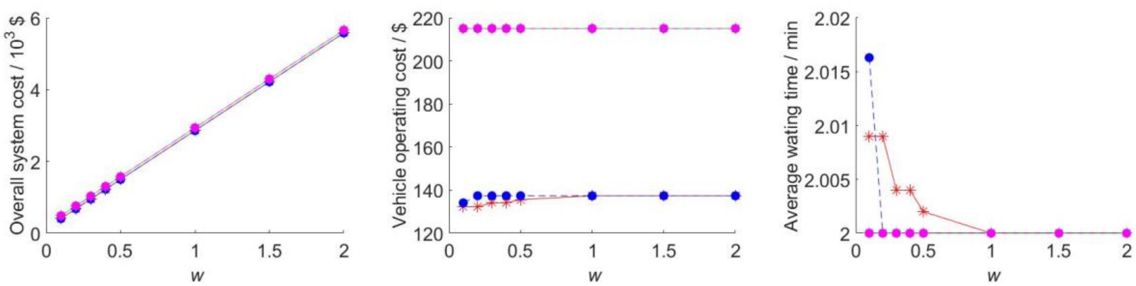
This subsection analyzes the effectiveness of the proposed operation with different input parameters. In each experiment, we vary one input parameter and keep the other parameters the same, as shown in the default setting of Tab. 5. We use the overall system cost, vehicle operating cost, and passenger average waiting time to evaluate the system performance. The results of (1) the proposed operation without considering the MV level limit (i.e., when each dispatched MV always has a sufficient capacity to take all passengers along its route), (2) the oversaturated benchmark operation (i.e., when all dispatched MVs have identical capacities that may not be sufficient to take all waiting passengers), and (3) the unsaturated benchmark operation (i.e., when all dispatched MVs have identical capacities always sufficient to take all waiting passengers) are also shown to compare with the results of the proposed operation. The results of these instances are plotted in Fig. 11.



(a) Performance with varying minimum headway



(b) Performance with varying the maximum number of the MUs allowed in an MV



(c) Performance with varying the value of waiting time

Fig. 11 Sensitivity analysis with different input parameters.

Fig. 11 (a) shows that as the minimum headway (i.e. H) increases, the effectiveness of the proposed operation weakens. This is because the waiting passengers at each station also increase as the minimum headway increases. To minimize the overall system cost, more MUs are dispatched for each MV, which reduces the gap between the vehicle operating costs of the benchmark operations and of the proposed operation. Note that in the overall system cost and vehicle operating cost plots, the curve of the proposed operation, considering the MV level limit, is slightly lower than that of the proposed operation without considering the MV level limit. To satisfy that all waiting passengers can board at each dispatch, one more MU needs to be dispatched, even if only one passenger is left. Since the capacity is sufficient for the passenger demand at each station even when the minimum headway equals 10 minutes (i.e. the maximum number of waiting passengers at each station), there is no difference between the oversaturated and unsaturated benchmark operations, and the corresponding curves overlap with each other as shown in Fig. 11 (a).

Fig. 11 (b) implies that as the MV level (i.e. L) increases, the vehicle operating cost of the proposed operation increases initially and then remains constant, while the vehicle operating cost of the benchmark operation keeps increasing. When the MV level reaches the maximum level in the optimal solution, further increase of the MV level cannot improve the solution any more. Note that in the average waiting time plot, the unsaturated benchmark operation has the highest value initially. The proposed operation and oversaturated benchmark operation overlap with each other and have the 2nd highest value. The proposed operation without the MV limit has the lowest value. The unsaturated benchmark operation increases the dispatch headway since each dispatch now takes more passengers. When L=2, the optimal solutions of these two operations are the same, since the value of L is too small to take advantage of the variable MV levels.

Fig. 11 (c) shows that as the value of waiting time (i.e. w) increases, allowing for oversaturated queues results in less benefits. Note that in the vehicle operating cost plot, the curve of the proposed operation without the MV limit is higher than that of the proposed operation initially, until the two curves gradually begin to overlap each other. This is because of the increasing weight of the waiting time cost in the overall system cost due to the increasing value of waiting time. As a result, to minimize the overall system cost, more MVs are dispatched to decrease the waiting time cost, which will consequently decrease or even eliminate oversaturated queues throughout the operation period. This makes the operations, whether or not allowing for oversaturated queues, eventually identical.

5. Conclusion

This paper investigates the MODOC problem and formulates it into a MILP model with the proposed MSFIFO rule. To solve the proposed model, theoretical properties of the invesitgated model are analyzed. These theoretical properties lead to identificaiton of upper and lower bounds to a feasible solution. Based on these two bounds, the demand clearance constraint and the MSFIFO rule, a customized DP algorithm with much reduced state pace and dimensions is developed to obtain the exact optimal solution in an acceptable computational time. Two sets of numerical experiments are implemented to verify the effectiveness of the proposed model and algorithm. From the numerical results, we see that the proposed customized DP algorithm much outperforms a state-of-the-art commercial solver (Gurobi) in both solution quality and time. Thus, it can be applied to real-world operations of UTS systems (e.g., urban rail transit, bus transit and bus rapid transit) that demand a relatively efficient solution time. Further, compared with the benchmark operation, both the average waiting time and the overall system cost are largely reduced by the proposed operation. This indicates the proposed operation, if implemented, will help reduce UTS expenses while improving the service quality.

Future research can be conducted in a few directions. In this paper, a MV is assumed to be formed only at the origin while its level remains the same across all the downstream stations. In the future, station-based MV reformation may be considered to allow a MV change its level at each intermediate station. Further, we may consider en-route docking operations that allow a MV to vary its level on a segment while it is running. This study only considers a corridor, and it is interesting to extend it to a general network system. On the theoretical side, it is worth investigating more theoretical properties such as tighter bounds, which can further expedite the customized DP algorithm.

Acknowledgement

The authors especially thank the colleague Zhiwei Chen, Saeid Soleimaniamiri and all the other colleagues for the valuable constructive comments on this paper. This paper is supported by the National Science Foundation though Grants CNS #1638355 and CMMI #1932452.

Reference

Albrecht, T., 2009. Automated timetable design for demand-oriented service on suburban railways. Public Transp. 1, 5–20.

Barrena, E., Canca, D., Coelho, L.C., Laporte, G., 2014a. Single-line rail rapid transit timetabling under dynamic passenger demand. Transp. Res. Part B Methodol. 70, 134–150.

Barrena, E., Canca, D., Coelho, L.C., Laporte, G., 2014b. Exact formulations and algorithm for the train timetabling problem with dynamic demand. Comput. Oper. Res. 44, 66–74.

Cacchiani, V., Furini, F., Kidd, M.P., 2016. Approaches to a real-world Train Timetabling Problem in a railway node. Omega (United Kingdom) 58, 97–110.

Canca, D., Barrena, E., Algaba, E., Zarzo, A., 2014. Design and analysis of demand-adapted railway timetables. J. Adv. Transp. 48, 119–137.

Caros, N, 2019. Dynamic operations of a mobility service with en-route transfers (Master's Thesis). New York University, US.

Ceder, A. (Avi), 2011. Public-transport vehicle scheduling with multi vehicle type. Transp. Res. Part C Emerg. Technol. 19, 485–497.

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

Chen, Z., Li, X., Zhou, X., 2019b. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Continuous modeling method. Transp. Res. Part B Methodol.

Gao, Y., Kroon, L., Schmidt, M., Yang, L., 2016. Rescheduling a metro line in an over-crowded situation after disruptions. Transp. Res. Part B Methodol. 93, 425–449.

Guo, Q.W., Chow, J.Y.J., Schonfeld, P., 2018. Stochastic dynamic switching in fixed and flexible transit services as market entry-exit real options. Transp. Res. Part C Emerg. Technol. 94, 288–306.

Hassold, S., Ceder, A.A., 2014. Public transport vehicle scheduling featuring multiple vehicle types. Transp. Res. Part B Methodol. 67, 129–143.

Huang, Y., Yang, L., Tang, T., Cao, F., Gao, Z., 2016. Saving energy and improving service quality: Bicriteria train scheduling in urban rail transit systems. IEEE Trans. Intell. Transp. Syst. 17, 3364–3379.

Liu, T., (Avi) Ceder, A., 2017. Deficit function related to public transport: 50 year retrospective, new developments, and prospects. Transp. Res. Part B Methodol. 100, 1–19.

Mo, P., Yang, L., Wang, Y., Qi, J., 2019. A flexible metro train scheduling approach to minimize energy cost and passenger waiting time. Comput. Ind. Eng. 132, 412–432.

Niu, H., Zhou, X., 2013. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. Transp. Res. Part C Emerg. Technol. 36, 212–230.

Niu, H., Zhou, X., Gao, R., 2015. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints. Transp. Res. Part B Methodol. 76, 117–135.

Shang, P., Li, R., Liu, Z., Yang, L., Wang, Y., 2018. Equity-oriented skip-stopping schedule optimization in an oversaturated urban rail transit network. Transp. Res. Part C Emerg. Technol. 89, 321–343.

Shi, J., Yang, L., Yang, J., Gao, Z., 2018. Service-oriented train timetabling with collaborative passenger flow control on an oversaturated metro line: An integer linear optimization approach. Transp. Res. Part B Methodol. 110, 26–59.

Yang, L., Qi, J., Li, S., Gao, Y., 2016. Collaborative optimization for train scheduling and train stop planning on high-speed railways. Omega (United Kingdom) 64, 57–76.

Yin, J., Tang, T., Yang, L., Gao, Z., Ran, B., 2016. Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: An approximate dynamic programming approach. Transp. Res. Part B Methodol. 91, 178–210.

Yin, J., Yang, L., Tang, T., Gao, Z., Ran, B., 2017. Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. Transp. Res. Part B Methodol. 97, 182–213.

Zhao, X., Hou, Z., Chen, J., Zhang, Y., Sun, J., 2018. Urban Rail Transit Scheduling under Time-Varying Passenger Demand. J. Adv. Transp. 2018, 1–9.

Sun, L., Jin, J.G., Lee, D.H., Axhausen, K.W., Erath, A., 2014. Demand-driven timetable design for metro services. Transp. Res. Part C Emerg. Technol. 46, 284–299.