Transportation Research Part B 126 (2019) 68–86

Contents lists available at ScienceDirect

Transportation Research Part B

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

Multiline Bus Bunching Control via Vehicle Substitution

Antoine Petit, Chao Lei, Yanfeng Ouyang*

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 2 August 2018
Revised 13 April 2019
Accepted 20 May 2019
Available online 6 June 2019

Keywords:
Bus bunching
Transit reliability
Approximate dynamic programming
Multiline control

ABSTRACT

Traditional bus bunching control methods (e.g., adding slack to schedules, adapting cruising speed), in one way or another, trade commercial speed for better system stability and, as a result, may impose the burden of additional travel time on passengers. Recently, a dynamic bus substitution strategy, where standby buses are dispatched to take over service from late/early buses, was proposed as an attempt to enhance system reliability without sacrificing too much passenger experience. This paper further studies this substitution strategy in the context of multiple bus lines under either time-independent or time-varying settings. In the latter scenario, the fleet of standby buses can be dynamically utilized to save on opportunity costs. We model the agency's substitution decisions and retired bus repositioning decisions as a stochastic dynamic program so as to obtain the optimal policy that minimizes the system-wide costs. Numerical results show that the dynamic substitution strategy can benefit from the "economies of scale" by pooling the standby fleet across lines, and there are also benefits from dynamic fleet management when transit demand varies over time. Numerical examples are presented to illustrate the applicability and advantage of the proposed strategy. The substitution strategy not only holds the promise to outperform traditional holding methods in terms of reducing passenger costs, they also can be used to complement other methods to better control very unstable systems.

1. Introduction

Bus systems are typically subject to random disturbances during operations. Those disturbances can be either minor (e.g. drivers' vagaries, special passenger needs) or severe (e.g. vehicle breakdown, traffic congestion). Even small disturbances when left uncontrolled can grow rapidly into large schedule deviations that appear equivalent to disruptions to the passengers (Newell and Potts, 1964). An almost inevitable consequence is that bus headways and spacings become so irregular that they would eventually end up "bunching" together (Osuna and Newell, 1972; Daganzo, 2008). This phenomenon is considered a plague in urban bus systems not only because it wreaks havoc on the punctuality and regularity of bus schedules but, more importantly, because it imposes a heavy burden on the passengers' travel experience (e.g., long waiting/dwelling times, crowded buses), which may in turn affect ridership and revenue.

Intervention strategies against small disturbances have been proposed in the last few decades; see Ibarra-Rojas et al. (2015) for a comprehensive review. Early methods consist of adding slacks into the schedule so as to retain early buses at certain stations and prevent the propagation of schedule discrepancies (Newell, 1974; Abkowitz and Lepofsky, 1990; Eberlein et al., 2001). Recently, other holding-based strategies were proposed to take advantage of real-time information and adaptively control the holding time or the speed of buses (Daganzo, 2009; Daganzo and Pilachowski, 2011; Xuan et al., 2011;

* Corresponding author. Tel: +1 217 333 9858
E-mail address: yfouyang@illinois.edu (Y. Ouyang).

https://doi.org/10.1016/j.trb.2019.05.009
0191-2615/

Bartholdi and Eisenstein, 2012; Berrebi et al., 2015; Sánchez-Martínez et al., 2016). Those strategies were designed to stabilize headways while minimizing passenger waiting time, typically for systems that operate without a schedule. Xuan et al. (2011) and a number of follow-up studies from the same authors adapted those methods and applied them to schedule-based systems. Such holding approaches generally (slightly) increase travel times of passengers for better stability of arrival times. In addition, it shall be noted that the effectiveness of these strategies would be compromised if deviations to the schedule exceed certain bounds (Daganzo and Ouyang, 2019). More drastic measures such as stop-skipping (Fu et al., 2003; Sun and Hickman, 2005; Liu et al., 2013), limited-boarding (Delgado et al., 2009; 2012) and short-turning (Ceder, 1990) are no panacea either, as they create frustration among the passengers who either have to walk to their stop, or wait for another bus. Alternative strategies such as transit signal priority (Liu et al., 2003; Ling and Shalaby, 2004; Estrada et al., 2016) and overtaking (Wu et al., 2017) could be challenging to implement as they depend highly on agencies' administration power and the city's roadway infrastructures. More recently, Petit et al. (2018) proposed a so-called bus *substitution* strategy, where the transit agency dispatches *standby* buses to take over service from late/early in-operation buses. Those late/early buses then display a "not-in-service" (NIS) sign, but still operate in a drop-off only mode to deliver all the remaining onboard passengers (before returning to the standby bus pool). They showed with one homogeneous bus route that the strategy holds the promise to outperform the traditional fixed slack method and compete with the most advanced speed control methods under certain circumstances (under either minor disturbances or severe disruptions).

While most bunching mitigation literature has focused on a single transit line, efforts on multiline transit systems are relatively rare, and existing studies mainly address traditional control strategies. For example, Hernández et al. (2015) extended the work of Delgado et al. (2012) on limited-boarding and holding strategies to a system with multiple bus lines. Argote-Cabanero et al. (2015) studied the adaptive control from Xuan et al. (2011) for multiline networks and presented a real-world application. Nevertheless, the potential benefits of applying the bus substitution strategy to multiline systems, either under homogeneous settings or under time-varying settings, have not been explored yet. For instance, rather than assuming arbitrarily located standby buses along a homogeneous transit route, as in Petit et al. (2018), pre-positioning of the standby buses could shorten the dispatching lead times for substituting buses if these buses are already located within the vicinity where delay is likely to occur in the near future. Moreover, when transit demand varies over time, it shall be beneficial to allow dynamic utilization of buses either for regular service or for substitution (as standby buses). Systematically addressing these opportunities would allow a more practical and effective bunching intervention strategy to be developed for the transit agencies.

In light of these needs, this paper develops a modeling framework to study how the bus substitution strategy can be applied to multiline transit systems, where standby buses are shared across lines and can be called upon to provide regular service (if needed). Under time-independent settings, the major decisions of the transit agency include two parts: (i) the bus substitution decision, i.e., how to dispatch a shared pool of standby buses to substitute operating buses across multiple lines; and (ii) the retired bus repositioning decision, i.e., where to reposition the retired buses over the service region such that they can be effectively utilized in the future. The overall objective is to deliver the best service, e.g., least waiting and dwelling times of passengers, while minimizing the total operating costs of the transit agency. A stochastic dynamic program in an infinite horizon is developed to reveal the optimal decision-making policy in a time-independent setting, where the bus substitution and prepositioning decisions, as well as the dynamics of bus operations across multiple lines are addressed. Approximate dynamic programming (ADP) is used with an embedded heuristic subroutine to solve the Bellman subproblem. Under time-varying settings, we divide the time horizon into multiple homogeneous periods, and allow dynamic utilization of buses (either to provide regular service or serve as standby) as part of the fleet management decision. A multi-period optimization program is proposed to determine the optimal dynamic fleet allocation in each period. The proposed strategies and solution algorithms are tested with a set of hypothetical transit networks, and compared with the existing speed control methods under homogeneous settings. It is shown that the dynamic substitution strategy can benefit from "economies of scale" by pooling standby buses across multiple lines, and outperform its counterparts in terms of decreasing passenger costs. We also quantify the benefit of dynamic fleet management under time-varying demand. Results indicate that optimal fleet utilization allows further savings on vehicle opportunity costs. A case study based on multiple lines of the Champaign-Urbana Mass Transit District is also presented to draw managerial insights. Finally, we investigate the feasibility of a hybrid approach where dynamic bus speed control (suitable for small schedule perturbations) and bus substitution strategies (suitable for large perturbations) are combined together to stabilize bus operations.

The remainder of the paper is organized as follows. Section 2 introduces the characteristics and operations of multiline transit systems as well as the mathematical formulation of the bus substitution problem under time-independent settings. Section 3 describes the solution algorithm based on ADP. Section 4 briefly explains how the strategy can be applied as a building block to cases in a time-varying setting. Then, Section 5 presents a series of numerical examples that collectively illustrate the performance and applicability of the proposed strategy. Finally, Section 6 provides concluding remarks and future research directions.

2. Time-independent multiline substitution

In this section, we introduce the basic modeling framework for dynamic bus substitution when the system settings are time-independent. We present the system dynamics and the substitution policy constraints, and propose the optimization model.

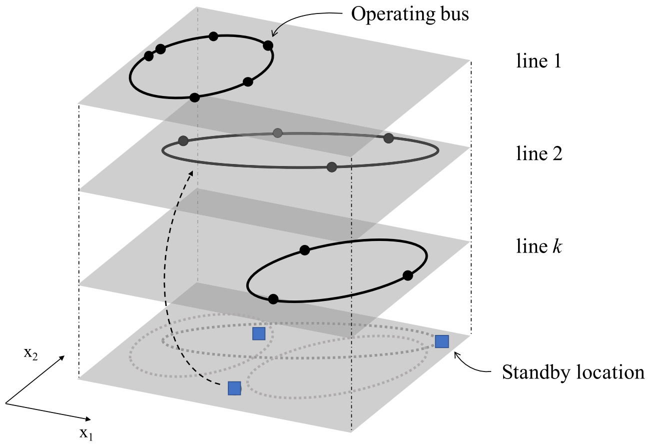

Fig. 1. Illustration of multiline transit system.

2.1. Model setting

Consider an urban transit system with a set of bus lines, denoted by $\mathcal{K}$, being distributed over a two-dimensional space, as shown in Fig. 1. Even though interdependencies may exist among different lines in the real world (e.g., due to passenger transfers and sharing of stations), here we assume they are independent of each other for the sake of simplicity. Similar to Petit et al. (2018), each single line $k \in \mathcal{K}$ is assumed to be operated along a corridor and treated as a bidirectional closed loop with length L_k. Without loss of generality, we only consider one direction of each of the loops. Let $\mathcal{N}_k$ denote the set of operating buses along one direction of bus line $k \in \mathcal{K}$, where each operating bus is indexed by tuple (k,n) for $k \in \mathcal{K}$ and $n \in \mathcal{N}_k$. It is assumed that all the buses in the system are identical and interchangeable, e.g., they have the same cruising speed, denoted by E, the same dwell time per passenger for alighting and boarding, denoted by A and B, respectively, and an infinite capacity.[1]

When bus lines are independent of each other, we take one bus line $k \in \mathcal{K}$ as a representative to illustrate the dynamics of bus operations. Assume that the passengers taking line $k \in \mathcal{K}$ arrive randomly with a steady rate over time and space,[2] denoted by λ_k, and all the buses are evenly spaced along the line, i.e., the average spacing between two consecutive buses S_k equals to $\frac{L_k}{|\mathcal{N}_k|}$. According to Petit et al. (2018), the "scheduled" commercial speed of buses in line k, denoted by V_k, can be defined as

$$V_k = E(1 - B\lambda_k S_k), \ \forall k \in \mathcal{K}. \tag{1}$$

It can be obtained by expressing the time traveling $(1/V_k)$ as the sum of the time cruising $(1/E)$, and the time dwelling $(B\lambda S_k/V_k)$, which depends on the number of passengers boarding per unit distance $(\lambda S_k/V_k)$. Note that we ignore the dwell time due to alighting.[3]

The dynamics of the "scheduled" trajectory for bus (k,n), denoted by $\{Y_{k,n}(t)\}_{t\geq 0}$, can be presented as

$$\begin{aligned} Y_{k,n}(t+1) &= Y_{k,n}(t) + V_k, \ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0, 1, 2, \dots\}, \\ Y_{k,n}(0) &= (|\mathcal{N}_k| - n) \cdot S_k, \ \forall k \in \mathcal{K}, n \in \mathcal{N}_k. \end{aligned} \tag{2}$$

However, as mentioned earlier, buses rarely stick to their schedules if they are left uncontrolled. Let $\tilde{\omega}_{k,n}(t)$ denote an additive disturbance to bus (k,n) at time t, which follows an independent and identically distributed (i.i.d.) normal distribution with mean 0 and variance σ_k^2. The uncontrolled trajectory of bus (k,n), denoted by $y_{k,n}(t)$, can then be written out similarly as in Pilachowski (2009):

$$y_{k,n}(t+1) = y_{k,n}(t) + v_{k,n}(t) + \tilde{\omega}_{k,n}(t), \ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0, 1, 2, \dots\}, \tag{3}$$

[1] Simulation results of a real-world case study in Section 5.3 suggest that, even in the cases of large schedule deviations, vehicle occupancies remain at reasonable levels and bus capacity does not become a limitation before substitution occurs. The readers can refer to Section 5.5 for more discussion on this assumption.

[2] This assumption is generally conservative for the proposed substitution strategy. First, Appendix B analytically shows that the passenger costs in this continuous model are slightly overestimated as compared to those with discrete bus stops. Appendix B also discusses the impact of the discrete stops on the frequency of bunching occurrences, and thus the frequency and cost of substitutions. Second, for the substitution strategy, it is actually more difficult to handle spatially homogeneous demand. As travel demand becomes unevenly distributed over space, the locations of bunching (and substitutions) are more likely to concentrate in certain neighborhoods. Then, standby vehicles could be positioned strategically so as to accelerate the insertion process and reduce the substitution costs.

[3] This is not a critical assumption. The proposed framework can also accommodate other real-world scenarios, e.g., when boarding and alighting occur sequentially such that the dwell time is measured by $A + B$.

where $v_{k,n}(t)$ denotes the instantaneous commercial speed of bus n at time t. Note that $v_{k,n}(t)$ can be derived by simply replacing $S_{k,n}$ in Eq. (1) with $s_{k,n}(t)$, which denotes the "actual" spacing of operating bus (k, n) at time t. Derivations of $s_{k,n}(t)$ can be found in Petit et al. (2018). We further assume that buses are not allowed to leapfrog[4], i.e.,

$$s_{k,n}(t) \geq 0, \ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0, 1, 2, \dots\}. \tag{4}$$

The deviation of bus (k,n) from schedule at time t, denoted by $\epsilon_{k,n}(t)$, can be obtained by calculating the difference between its "scheduled" and "actual" positions at that moment, i.e.,

$$\epsilon_{k,n}(t) = Y_{k,n}(t) - y_{k,n}(t), \ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0, 1, 2, \dots\}, \tag{5}$$

where $\epsilon_{k,n}(t) > 0$ ($\epsilon_{k,n}(t) < 0$) indicates that the bus is late (early). Meanwhile, the actual bus occupancy, denoted by $O_{k,n}(t)$, can be derived based on Eq. (9) in Petit et al. (2018):

$$O_{k,n}(t+1) \approx \max\left\{O_{k,n}(t)\left[1 - \frac{4v_{k,n}(t)}{L_k}\left(1 - \frac{v_{k,n}(t)}{L_k}\right)\right], 0\right\} + \lambda_k s_{k,n}(t), \ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0, 1, 2, \dots\}. \tag{6}$$

The first term in Eq. (6) pertains to the number of the alighting passengers, which is based on the triangular distribution of the destinations of the onboard passengers (see Appendix A in Petit et al. (2018)). It is assumed that passengers travel at most a distance of $L/2$. The second term pertains to the number of boarding passengers, which is proportional to the spacing in front of the bus.

2.2. Bus substitution strategy

As shown in Fig. 1, standby buses are prepositioned at certain locations that are close to the bus routes (e.g., nearby parking lots). Denoting $\mathcal{J}$ as the set of parking locations and $w_j(t)$ as the number of standby buses that are held at $j \in \mathcal{J}$ at time t, the transit agency's primary decision is then to dynamically determine whether a substitution should be made such that the system performance can be enhanced. Let $\zeta_{j,k,n}(t) \in \{0, 1\}, \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k$ denote the agency's substitution decision at t, where $\zeta_{j,k,n}(t) = 1$ if a standby bus located at j is dispatched to substitute operating bus (k, n); or $\zeta_{j,k,n}(t) = 0$ otherwise.

Due to the limited size of the standby bus fleet, the bus substitution decision at time t should be restricted by the number of available buses at the parking locations, i.e.,

$$\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}_k} \zeta_{j,k,n}(t) \leq w_j(t), \ \forall j \in \mathcal{J}, t \in \{0, 1, 2, \dots\}. \tag{7}$$

Meanwhile, an operating bus can be substituted with at most one standby bus, i.e.,

$$\sum_{j\in\mathcal{J}} \zeta_{j,k,n}(t) \leq 1, \ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0, 1, 2, \dots\}, \tag{8}$$

and we further assume that the substitution cannot be applied to an operating bus that has just been inserted into operation recently (i.e., whose corresponding retiring bus is still running in NIS mode). Let $y^R_{k,n}(t)$ denote the position of NIS bus (k, n) at time t, where $y^R_{k,n}(t) > 0$ if bus (k, n) was just substituted and is currently going through the retirement process, while $y^R_{k,n}(t) < 0$ if the NIS mode has not been activated yet. Thus, the following constraints should be applied to the substitution decision:

$$\sum_{j\in\mathcal{J}} \zeta_{j,k,n}(t) \leq 1 - \frac{y^R_{k,n}(t) \pmod{L_k}}{L_k}, \ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0, 1, 2, \dots\}, \tag{9}$$

where *mod* refers to the modulo operation, where the remainder can be negative as defined in Knuth (1972).

Due to the time needed for moving a standby bus from its parking location to the target insertion position along the line, the lead time between the substitution "decision" and actual "action" needs to be explicitly considered. For simplicity of modeling, we assume that the substitution can only happen if the associated lead time is no longer than the time interval between two consecutive decision epochs, denoted by Δ, i.e.,

$$\zeta_{j,k,n}(t) \leq 1 - \frac{T[j, Y_{k,n}(t+\Delta) \pmod{L_k}] - \Delta}{T_{\max}}, \ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0, 1, 2, \dots\}, \tag{10}$$

where $T_{\max}$ denotes the maximum travel time between any two points in the studied area and $T[j, Y_{k,n}(t+\Delta) \pmod{L_k}]$ denotes the time needed to travel from parking location $j \in \mathcal{J}$ to the scheduled location of bus (k,n) at time $(t+\Delta)$, denoted by $Y_{k,n}(t+\Delta) \pmod{L_k}$.

Fig. 2 gives a simple illustration of a bus trajectory involving a substitution for a late operating bus. It can be observed that the late bus is reset right back to its schedule at time $t+\Delta$, after a substitution decision has been made at t. Hence,

[4] To be conservative, we assume that the agency does not allow overtaking, e.g., for safety reasons. Such an assumption is not too restrictive, since we only need to swap the involved bus indices when leapfrog occurs, while keeping the spacing non-negative.

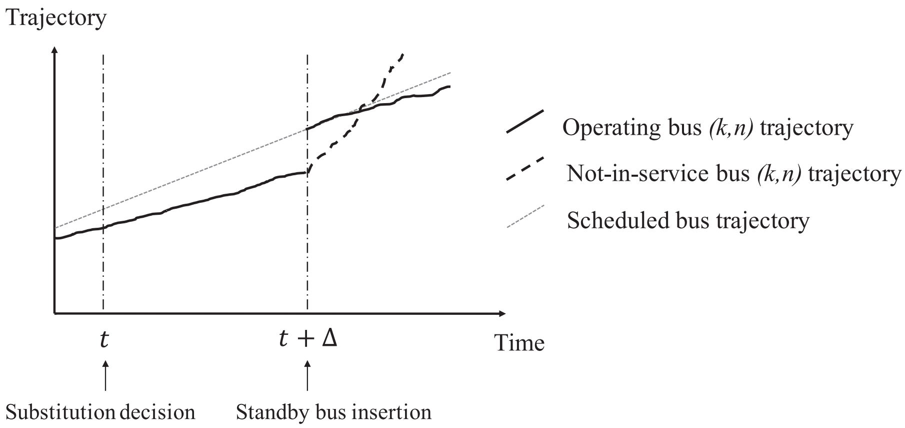

Fig. 2. Bus trajectory with substitution (from Petit et al. (2018)).

when the substitution strategy is applied, the bus deviations from the schedule can be formulated as follows:

$$\epsilon_{k,n}(t) = \left(1 - \sum_{j\in\mathcal{J}} \bar{\zeta}_{j,k,n}(t-\Delta)\right) \cdot \left(Y_{k,n}(t) - y_{k,n}(t)\right),\ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{\Delta, \Delta+1, \ldots\}, \tag{11}$$

where $\bar{\zeta}_{j,k,n}(t-\Delta)$ denotes the substitution decision that was made at time $t-\Delta$, which is already known at time t.

Moreover, if a substitution decision has been made for bus (k,n), i.e., $\zeta_{j,k,n}(t) = 1$, the inserted bus is then denoted by the same index (k,n) starting from time $t+\Delta$. In other words, operating bus (k,n) is treated as remaining in the system but with an instantaneous location change upon substitution.

2.3. Not-in-service bus retirement and repositioning

During substitution, an operating bus would retire and change into not-in-service (NIS) mode. In this mode, a bus only allows onboard passengers to alight, until it is empty. Once empty, it needs to be repositioned at a parking location. We assume that the transit agency decides at the time of the substitution which parking location this NIS bus should be sent to. Let $\phi_{k,n,j}(t) \in \{0,1\}, k \in \mathcal{K}, n \in \mathcal{N}_k, \forall j \in \mathcal{J}$ denote the agency's repositioning decision for NIS buses, where $\phi_{k,n,j}(t) = 1$ if NIS bus (k,n) would be sent to parking location $j \in \mathcal{J}$, and $\phi_{k,n,j}(t) = 0$ otherwise. Since a bus that has been substituted must be repositionned, the following equations hold:

$$\sum_{j\in\mathcal{J}} \phi_{k,n,j}(t) = \sum_{j\in\mathcal{J}} \bar{\zeta}_{j,k,n}(t-\Delta),\ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{\Delta, \Delta+1, \ldots\}. \tag{12}$$

After the repositioning decision has been made at t, assuming that $\phi_{k,n,j}(t) = 1$, the NIS bus would go through two phases thereafter: (i) dropping off onboard passengers until it becomes empty, and (ii) traveling to the designated parking location as directly as possible. In phase (i), it is assumed that the NIS bus travels a distance of $L_k/2$, which is the upper bound of the distance needed to drop off all its onboard passengers. The trajectory of NIS bus (k,n) along the line can be derived as[5]

$$y^R_{k,n}(t'+1) = y^R_{k,n}(t') + v^R_{k,n}(t'),\ \forall t' \in \{t, t+1, \ldots\}, \tag{13}$$

where $v^R_{k,n}$ denotes its instantaneous commercial speed defined as

$$v^R_{k,n}(t') \approx \begin{cases} \left[E^{-1} + \max\left\{\frac{A}{L_k/2 + y^R_{k,n}(t) - y^R_{k,n}(t')}\left(2 - \frac{1}{L_k/2 + y^R_{k,n}(t) - y^R_{k,n}(t')}\right) O^R_{k,n}(t'), 0\right\}\right]^{-1} & \text{if } O^R_{k,n}(t') > 0, \\ 0 & \text{otherwise,} \end{cases}$$
$$\forall t' \in \{t, t+1, \ldots\}. \tag{14}$$

The first term pertains to the cruising time. The second term pertains to the dwell time due to passengers alighting. Note that the number of passengers alighting from the NIS buses is derived similarly to Eq. (6), except that the maximum distance that the onboard passengers can travel now reduces while the NIS bus advances.

[5] Similar to Eq. (10) in Petit et al. (2018), the random disturbance term is omitted.

The occupancy of the NIS buses, $O^R_{k,n}$, is given as

$$O^R_{k,n}(t'+1) \approx \max\left\{O^R_{k,n}(t')\left[1-\frac{v^R_{k,n}(t')}{L_k/2+y^R_{k,n}(t)-y^R_{k,n}(t')}\left(2-\frac{v^R_{k,n}(t')}{L_k/2+y^R_{k,n}(t)-y^R_{k,n}(t')}\right)\right],0\right\},$$
$$\forall t' \in \{t, t+1, \dots\}. \quad (15)$$

Eq. (15) is derived similarly to Eq. (6), except that the maximum distance that the retiring bus will travel while dropping off passengers is now reduced. If $y^R_{k,n}(t)$ is the position of the retiring bus at the time it starts operating as NIS, and $y^R_{k,n}(t')$ is the current position of the retiring bus, the distance left to travel is $L_k/2-(y^R_{k,n}(t')-y^R_{k,n}(t))$.

In phase (ii), the empty NIS bus (k,n) takes the shortest path to reach the designated parking location. To keep track of the status of NIS bus (k,n) during the whole retirement process, we calculate at time t the remaining time needed to reach parking location j, $\tau^R_{k,n,j}(t)$, as follows:

$$\tau^R_{k,n,j}(t) = \max\left\{0, \phi_{k,n,j}(t)\cdot\left(A\cdot O^R_{k,n}(t)+\frac{L_k}{2E}+T\left[\left(y^R_{k,n}(t)+L_k/2\right) \pmod{L_k}, j\right]\right)\right.$$
$$\left.+\left(1-\phi_{k,n,j}(t)\right)\cdot\left(\tau^R_{k,n,j}(t-\Delta)-\Delta\right)\right\}, \ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\}, \quad (16)$$

where $A\cdot O^R_{k,n}(t)$, $\frac{L_k}{2E}$ and $T\left[\left(y^R_{k,n}(t)+L_k/2\right) \pmod{L_k}, j\right]$ represent the time for dropping off all onboard passengers, traveling half the length of the line, and moving the empty bus to parking location j, respectively. The summation of these three terms corresponds to the remaining time for bus (k,n) to reach parking lot j if it is substituted at time t, i.e., $\phi_{k,n,j}(t)=1$. Otherwise, simply deduct Δ from the remaining time at time $t-\Delta$, i.e., $\tau^R_{k,n,j}(t-\Delta)-\Delta$. Then, we let $u^R_{k,n,j}(t) \in \{0,1\}, \forall k \in \mathcal{K}, n \in \mathcal{N}_k, j \in \mathcal{J}$ denote whether or not NIS bus (k, n) reaches parking location j before the next decision epoch, i.e., $u^R_{k,n,j}(t)=1$ if NIS bus (k,n) is going to reach parking location j before $t+\Delta$, and $u^R_{k,n,j}(t)=0$ otherwise. The relationship between $u^R_{k,n,j}(t)$ and $\tau^R_{k,n,j}(t)$ is described in Appendix A.

The transition functions for NIS bus trajectories can be expressed as

$$y^R_{k,n}(t) = \left(\sum_{j\in\mathcal{J}}\bar{\zeta}_{j,k,n}(t-\Delta)\right)\cdot y_{k,n}(t) - \left(1-\sum_{j\in\mathcal{J}}\bar{\zeta}_{j,k,n}(t-\Delta)\right)\left[\sum_{j\in\mathcal{J}}u^R_{k,n,j}(t-\Delta)\cdot y_{k,n}(t)\right.$$
$$\left.-\left(1-\sum_{j\in\mathcal{J}}u^R_{k,n,j}(t-\Delta)\right)\cdot y^R_{k,n}(t)\right], \ \forall k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{\Delta, \Delta+1, \dots\}. \quad (17)$$

The first term on the right-hand side of Eq. (17) corresponds to the initialization of the NIS bus trajectory when the operating bus switches to NIS mode, while the second term indicates that the trajectory of the NIS bus is reset to $-y_{k,n}(t)$[6] if the associated retirement process is completed between time t and $t+\Delta$, i.e., the NIS bus reaches its designated parking location.

Jointly considering the substitution decisions and the retired buses repositioning decisions, the conservation of standby resources over time can be written as

$$w_j(t+\Delta) = w_j(t) - \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}_k}\zeta_{j,k,n}(t) + \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}_k}u^R_{k,n,j}(t), \ \forall j \in \mathcal{J}, t \in \{0,1,2,\dots\}. \quad (18)$$

2.4. Objective

The transit agency aims at minimizing the total passengers' costs and substitution costs over the infinite horizon. Passengers costs pertain to the average waiting times before boarding the bus, as well as the delays that they experience when they alight from the buses (as compared to the published schedule). Similar to Eq. (17) in Petit et al. (2018), the passenger costs for using operating buses, denoted by $C^{\text{pax}}_{k,n}(t)$, can be presented as follows:

$$C^{\text{pax}}_{k,n}(t) = \mu\left[\left(\rho^{\text{late}}\max\left\{\frac{\epsilon_{k,n}(t)}{V_k},0\right\} - \rho^{\text{early}}\min\left\{\frac{\epsilon_{k,n}(t)}{V_k},0\right\}\right)\cdot\frac{4O_{k,n}(t)v_{k,n}(t)}{L_k}\left(1-\frac{v_{k,n}(t)}{L_k}\right)\right.$$
$$\left.+\frac{1}{2}\frac{\lambda_k s^2_{k,n}(t)}{v_{k,n}(t)}\right], \forall t, k \in \mathcal{K}, n \in \mathcal{N}_k, \quad (19)$$

where μ denotes the passenger's value of time, ρ^{late} is the weight associated with lateness, and ρ^{early} is the weight associated with earliness.

[6] This value is arbitrary. Using a negative value allows us to simplify Eq. (9).

The substitution costs include the costs for (i) requisitioning standby buses, (ii) operating NIS buses until they are empty and repositioning them, as well as (iii) the costs associated with the passengers alighting from NIS buses, which can be formulated as follows:[7]

$$C^R_{j,k,n}(t) = \Big\{c_M \cdot T\big[j, Y_{k,n}(t+\Delta) \pmod{L_k}\big]\Big\} \cdot \zeta_{j,k,n}(t) + \left\{c_M \cdot \tau^R_{k,n,j}(t) + \int_t^{t+\tau^R_{k,n,j}(t)} C^{R,\text{pax}}_{k,n}(t')dt'\right\} \cdot \phi_{k,n,j}(t), \forall t, k \in \mathcal{K}, n \in \mathcal{N}_k, j \in \mathcal{J}, \tag{20}$$

where c_M denotes the agency operating cost per unit of time, and $C^{R,\text{pax}}_{k,n}(t')$ denotes the passenger costs incurred on NIS buses, i.e.,

$$C^{R,\text{pax}}_{k,n}(t') = \mu\left(\rho^{\text{late}} \max\left\{\frac{\epsilon^R_{k,n}(t')}{V_k}, 0\right\} - \rho^{\text{early}} \min\left\{\frac{\epsilon^R_{k,n}(t')}{V_k}, 0\right\}\right) \cdot \max\left\{\frac{O^R_{k,n}(t')v^R_{k,n}(t')}{L_k/2 + \epsilon^R_{k,n}(t') - Y_{k,n}(t') + y^R_{k,n}(t)} \cdot \left(2 - \frac{v^R_{k,n}(t')}{L_k/2 + \epsilon^R_{k,n}(t') - Y_{k,n}(t') + y^R_{k,n}(t)}\right), 0\right\}, \forall t' \geq t, k \in \mathcal{K}, n \in \mathcal{N}_k. \tag{21}$$

Note that there are no waiting costs associated with NIS buses in Eq. (21) as they are drop-off only.

At this point, the dynamic multiline bus substitution problem over the infinite horizon can be formulated as follows:

$$\min_{\zeta,\phi} \sum_{t=0}^{\infty} \sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}_k} \left\{\mathbb{E}\left[\int_t^{t+\Delta} C^{\text{pax}}_{k,n}(t)dt\right] + \sum_{j\in\mathcal{J}} C^R_{j,k,n}(t)\right\} \tag{22}$$

$$\text{s.t. } (3)-(21), \text{ and}$$
$$\zeta_{j,k,n}(t), \phi_{k,n,j}(t) \in \{0,1\}, \ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\ldots\}.$$

3. Solution approach

The state space of the infinite-horizon stochastic dynamic program (22), denoted by Ω, consists of the following elements: the deviations of operating buses from their schedules $\boldsymbol{\epsilon} = \{\epsilon_{k,n}\}_{k\in\mathcal{K},n\in\mathcal{N}}$, their occupancies $\boldsymbol{O} = \{O_{k,n}\}_{k\in\mathcal{K},n\in\mathcal{N}}$, the status of NIS buses $\boldsymbol{\tau^R} = \{\tau^R_{k,n,j}\}_{k\in\mathcal{K},n\in\mathcal{N}_k,j\in\mathcal{J}}$, their trajectory $\boldsymbol{y^R} = \{y^R_{k,n}\}_{k\in\mathcal{K},n\in\mathcal{N}_k}$, and their occupancy $\boldsymbol{O^R} = \{O^R_{k,n}\}_{k\in\mathcal{K},n\in\mathcal{N}_k}$, and the number of standby buses at different parking locations $\boldsymbol{w} = \{w_j\}_{j\in\mathcal{J}}$. As such, the value function of the state ($\boldsymbol{\epsilon}, \boldsymbol{O}, \boldsymbol{\tau^R}, \boldsymbol{y^R}, \boldsymbol{O^R}, \boldsymbol{w}$) should satisfy the Bellman's optimality equation:

$$J(\boldsymbol{\epsilon}, \boldsymbol{O}, \boldsymbol{\tau^R}, \boldsymbol{y^R}, \boldsymbol{O^R}, \boldsymbol{w}) = \min_{\zeta,\phi} \sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}_k} \sum_{j\in\mathcal{J}} C^R_{j,k,n}(t) + \mathbb{E}\left[\sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}_k} \int_t^{t+\Delta} C^{\text{pax}}_{k,n}(t')dt' + J\left(\hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{O}}, \hat{\boldsymbol{\tau}}^{\boldsymbol{R}}, \hat{\boldsymbol{y}}^{\boldsymbol{R}}, \hat{\boldsymbol{O}}^{\boldsymbol{R}}, \hat{\boldsymbol{w}}\right)\right], \forall\left(\boldsymbol{\epsilon}, \boldsymbol{O}, \boldsymbol{\tau^R}, \boldsymbol{y^R}, \boldsymbol{O^R}, \boldsymbol{w}\right) \in \Omega, t \in \{0,1,2,\ldots\}, \tag{23}$$

where $\left(\hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{O}}, \hat{\boldsymbol{\tau}}^{\boldsymbol{R}}, \hat{\boldsymbol{y}}^{\boldsymbol{R}}, \hat{\boldsymbol{O}}^{\boldsymbol{R}}, \hat{\boldsymbol{w}}\right) \in \Omega$ denotes the state that the system transitions to at time $t + \Delta$.

Given the huge size of the state space and the binary nature of the control decisions, we proposed to solve the model by using an ADP-based algorithm whose basic framework is presented in pseudocode in Algorithm 1. The procedures of the ADP algorithm are quite standard and can be found in the relevant literature (Powell, 2011; Petit et al., 2018).

For the modeling simplicity, we consider the value function approximation (VFA) as functions of state vectors $\boldsymbol{\epsilon}$ and $\boldsymbol{w}$. The occupancy vector $\boldsymbol{O}$ is ignored since occupancies are highly correlated with the deviations $\boldsymbol{\epsilon}$, while the status information of NIS buses, i.e., $\boldsymbol{\tau}^R$, is less important than the availability of those buses at each standby location, i.e., $\boldsymbol{w}$. Therefore, the VFA denoted by $\widetilde{J}(\boldsymbol{\epsilon}, \boldsymbol{w})$ can be defined as a finite set of one-dimensional separable functions as

$$\widetilde{J}(\boldsymbol{\epsilon}, \boldsymbol{w}) = \sum_{j\in\mathcal{J}} \psi_j\left(w_j\right) + \sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}_k} \psi_k\left(\epsilon_{k,n}\right),$$

where $\psi_j(w_j)$ and $\psi_k(\epsilon_{k,n})$ are the piecewise linear functions with respect to w_j and $\epsilon_{k,n}$. The piecewise linear functions are adopted since they are simple but meanwhile more stable than purely linear functions.

Given the state of the system, the term $\sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}_k} \int_t^{t+\Delta} C^{\text{pax}}_{k,n}(t')dt'$ in Eq. (24) is independent of the decisions $\boldsymbol{\zeta}$, $\boldsymbol{\phi}$. A Monte Carlo simulation method is then used to estimate its expected value. Those fixed-step simulations are based on Eqs. (3)–(6) and (13)–(16). The remaining combinatorial optimization problem involves the objective terms $\sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}_k} \sum_{j\in\mathcal{J}} C^R_{j,k,n}(t)$ and $\mathbb{E}\left[\widetilde{J}^{(r)}\left(\hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{w}}\right)\right]$. It can be solved with techniques such as enumeration or meta-heuristics (e.g., simulated annealing). Note that the second term is estimated using the final states generated by the Monte Carlo method.

[7] Standby buses could also be directly repositioned from one parking location to another for better utilization at later decision epochs. For simplicity, we chose to ignore this decision in the current study.

Algorithm 1 ADP algorithm

1: Generate a random set of states.
2: Initialize $\widetilde{J}$ for state variables $(\boldsymbol{\epsilon}, \boldsymbol{w})$.
3: **for** $r = 1$ to R **do**
4: Initialize the system state $\boldsymbol{\epsilon}^{(r)}, \boldsymbol{w}^{(r)}, \boldsymbol{O}^{(r)} = \{O_{eq}\}_{n \in N}, \boldsymbol{y}^{\boldsymbol{R}^{(r)}}, \boldsymbol{O}^{\boldsymbol{R}^{(r)}}, \boldsymbol{u}^{\boldsymbol{R}^{(r)}}$.
5: Sample the set of noise vectors $\left\{\{\boldsymbol{\omega}^{(r,q)}(t)\}\right\}_{q \in 1,\ldots,Q}$ for $t = 0, 1, \ldots, \Delta$.
6: **for** $q = 1$ to Q **do**
7: Solve the one-stage problem:

$$\min_{\zeta,\phi} \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}_k} \sum_{j \in \mathcal{J}} C^R_{j,k,n}(t) + \mathbb{E}\left[\sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}_k} \int_t^{t+\Delta} C^{\text{pax}}_{k,n}(t)dt + \widetilde{J}^{(r)}(\hat{\epsilon}, \hat{\omega})\right]. \quad (24)$$

8: Obtain the marginal information $\mathbf{v}^{(r,q)}$, and update the VFA.
9: Update system state using Eqs. (3)-(18), $\{\boldsymbol{\omega}^{(r,q)}(t)\}_{t=0,1,\ldots\Delta}$, and solution from the previous step.
10: **end for**
11: **end for**
12: **return** the current value function $\widetilde{J}$.

Numerical gradients of the state variables are used as the marginal value information $\mathbf{v}^{(r,q)}$ that are required for updating the VFA in step 8 of Algorithm 1. They are obtained by (i) slightly perturbing each element of the state variables independently, (ii) solving the corresponding one-stage problem (), and then (iii) calculating the differences in objective function values. Given the marginal value information $\mathbf{v}^{(r,q)}$, the VFA can be updated according to the following equation:

$$\widetilde{J}^{(r)}(\boldsymbol{\epsilon}, \boldsymbol{w}) = \begin{cases} \left(1 - \theta^{r-1}\right)\widetilde{J}^{(r-1)}(\boldsymbol{\epsilon}, \boldsymbol{w}) + \theta^{r-1}\mathbf{v}^{(r,q)}, & \text{for } (\boldsymbol{\epsilon}, \boldsymbol{w}) = \left(\boldsymbol{\epsilon}^{(r,q)}, \boldsymbol{w}^{(r,q)}\right), \\ \widetilde{J}^{(r-1)}(\boldsymbol{\epsilon}, \boldsymbol{w}), & \text{otherwise}, \end{cases}$$

where $\theta^{r-1} \in [0, 1]$ is the step size parameter (e.g., generalized harmonic step size).

The implementation of the proposed decision policy can be conducted in two phases: (i) the “offline” training process, where the VFA is improved continuously by letting the ADP algorithm run for an arbitrary length of time over a large static dataset (either historical or simulated); (ii) the “online” decision-making process, where the obtained VFA policy is implemented in real time by solving the one-stage subproblem (24) based on the current state of the system. In practice, as the decision intervals of the transit agency are usually on the order of 10 to 15 minutes, dispatchers can easily apply the proposed approach and make decisions in real time.

4. Time-varying problem

Transit agencies may have to deal with time-varying settings (e.g., surge in travel demand, time-dependent congestion) and need to adapt their service to best serve the passengers while keeping their operating budget. Thereby, the optimal fleet utilization between regular and standby services is likely to change from one time period to another. In this section, we briefly discuss the multi-period problem which aims at optimizing the transit fleet size and utilization over a non-homogeneous planning horizon.

Building upon the model from Section 2.1, we consider a planning horizon $[T_1, T_{P+1}]$ (e.g., a day) that can be divided into a set of P relatively homogeneous long time periods (e.g., peak hours, off-peak hours), denoted by $\mathcal{P}$. These time periods start at discrete times $\{T_1, T_2, \ldots T_P\}$. The duration of each of these periods (e.g., 4-8 hours) is long as compared to the decision epoch length (e.g., 10 min), and within each period the system parameters are assumed to be invariant.

In period $p \in \mathcal{P}$, the state of the system is now captured by the number of operating buses across all lines $\{N_{p,k}\}_{k \in \mathcal{K}}$, and the number of standby buses S_p. We denote the entire bus utilization plan by $\boldsymbol{N} = \{N_{p,k}\}_{k \in \mathcal{K}, p \in \mathcal{P}}$ and $\boldsymbol{S} = \{S_p\}_{p \in \mathcal{P}}$. We use c_{acq} to denote the prorated costs of acquiring a bus, as well as the associated depreciation costs, per unit time. The multi-period planning problem comes down to determining the optimal fleet size, and the utilization of these vehicles during each time period, and the optimal substitution policy $(\boldsymbol{\zeta}, \boldsymbol{\phi}) = \left(\{\zeta_{p,j,k,n}(t)\}, \{\phi_{p,j,k,n}(t)\}\right)$, where $\{\zeta_{p,j,k,n}(t)\}$, and $\{\phi_{p,j,k,n}(t)\}$ are the substitution and repositioning decisions during each period $p \in \mathcal{P}$. If we replace the terms $\mathcal{N}_k$, $\zeta_{j,k,n}(t)$, and $\phi_{j,k,n}(t)$ in Eq. (3)-(21) by their multi-period counterparts, $\mathcal{N}_{p,k}$, $\zeta_{p,j,k,n}(t)$, and $\phi_{p,j,k,n}(t)$, respectively, the multi-period version of the model can be formulated as follows:

$$\min_{\boldsymbol{N},\boldsymbol{S},\boldsymbol{\zeta},\boldsymbol{\phi}} \sum_{p=1}^{P} \int_{t=T_p}^{T_{p+1}} \left(\sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}_{p,k}} \left\{\mathbb{E}\left[\int_t^{t+\Delta} C^{\text{pax}}_{k,n}(t')dt'\right] + \sum_{j \in \mathcal{J}} C^R_{j,k,n}(t)\right\}\right) dt + c_{acq} \cdot \left(\sum_{k \in \mathcal{K}} N_{1,k} + S_1\right) \quad (25)$$

$$\text{s.t.} \; (3) - (21), \text{and}$$

$$\sum_{k\in\mathcal{K}} N_{p,k} + S_p = \sum_{k\in\mathcal{K}} N_{p-1,k} + S_{p-1},\ \forall p \in \mathcal{P}\backslash\{1\}, \tag{26}$$
$$N_{p,k} \in \mathbb{N},\ \forall k \in \mathcal{K}, \forall p \in \mathcal{P},$$
$$S_p \in \mathbb{N},\ \forall p \in \mathcal{P},$$
$$\zeta_{p,j,k,n}(t), \phi_{p,k,n,j}(t) \in \{0,1\},\ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\},\ \forall p \in \mathcal{P}.$$

The objective function now includes the passenger and agency operating costs across the entire time horizon, along with the agency's investment for acquiring the fleet (i.e., the second term in (25)). Constraints (26) state that fleet size decisions are at the tactical level; i.e., the total number of buses in the fleet shall be constant in the entire time horizon.

While solving problem (25), we assume that each time period is long enough so that the system will quickly reach its stationary state. As such, the infinite-horizon model and solution from Sections 2-3 can be directly applied to each of these periods. Conditional on the fleet configuration decisions (***N, S***), bus substitution decisions within each of the time periods are similar to the solution to Problem (22). Based on that, we propose a simple heuristic to determine ***N***, where all reasonable fleet combinations are tested for each time period, with the corresponding system costs obtained from the model in Section 2. To reduce the computation burden, for any given operating fleet, we only tested a few sizes of the standby bus fleet — we stop as soon as the marginal benefit of having one additional standby bus in the fleet becomes negligible. Recall that a bus could be unused (i.e., neither assigned as an operating bus nor used as a standby bus) in certain low demand periods. In such a case, this bus's depreciation costs could be lower.

In real-world settings, the transition of the bus fleet between the heterogeneous time periods could be made at the dispatchers' discretion. Since the change in real-world travel demand is not expected to occur sharply (in contrast to what is assumed in this section), dispatchers could progressively retire (or add) buses around the transition between time periods so as to eventually meet the target fleet sizes. They could, for instance, retire or add buses at the termini.

5. Numerical experiments

In this section, we conduct a set of experiments to test the performance of the proposed strategy in a variety of application scenarios. The purpose of the following experiments is to (i) show the benefits of pooling standby buses across an increasing number of transit lines, (ii) show how time-varying demand can be addressed with multi-period fleet management, (iii) illustrate how the proposed strategy can be applied to real-world systems, (iv) demonstrate the benefits of using bus substitution strategy as a complement to existing speed control strategies, and (v) investigate the impact of finite vehicle capacity.

Throughout this section, it is assumed that the agency makes the substitution decisions every $\Delta = 10$ min over (single or multiple) periods of 4 hours. A 30-sec time step is used to carry out the simulation of the transit system dynamics. All distances in this section are calculated based on L1 norm to best represent an underlying dense grid of streets. In each experiment, all standby buses are initially positioned uniformly across the standby locations.

All tests are performed on a desktop computer with 2.5 GHz CPU and 16 GB RAM. To stay conservative, here we consider that the early buses yield no delay cost if alighting passengers arrive early at their destinations, i.e., $\rho^{\text{late}} = 1$ and $\rho^{\text{early}} = 0$. Note that we also do not account for the benefits from early passenger arrivals.

5.1. Benefits of standby vehicles pooling

In this section, we present a set of experiments on a hypothetical transit network of different sizes to show the benefits of pooling the standby buses across multiple transit lines. For simplicity we consider only one demand period. The system costs are compared to those associated with the speed control strategies.

All the lines are set with the same characteristics: $L_k = 8$ km, $N_k = 4$ buses, $\lambda_k = 40$ pax/km-h, $\forall k \in \mathcal{K}$. The disturbances follow a normal distribution with the standard deviation $\sigma_k = 0.045$ km. Buses cruise at speed $E = 30$ km/h, passengers board at rate $B = 4$ s/pax and alight at rate $A = 2$ s/pax. In each scenario, we consider two parking locations, whose positions are fixed across all scenarios. Fig. 3 shows the route layout and parking locations for five scenarios, each with different numbers of bus lines and standby buses. The monetary cost of time for passengers is set to be $\mu = 20$ \$/h and the agency operating costs are $c_M = 37$ \$/h (Petit et al., 2018). For all scenarios, an additional fixed depreciation cost of 12 \$/h (Neff and Dickens, 2015) is also associated with each transit vehicle in the fleet, regardless of how it is operated.

Fig. 4 plots the resulting average costs for each scenario and compares those with the one-way looking strategy (Daganzo, 2009), the two-way looking strategy (Daganzo and Pilachowski, 2011), and the ideal system where there is no disturbance. We assume that the standby buses are unused for those three counterparts.[8] The costs for both the one-way looking and the two-way looking strategies are derived based on the reduced commercial speed proposed in the corresponding references. Since the travel demand is homogeneous along each line, it is assumed that all passengers travel on average a distance of $L/4$. We assume favorably that passengers do not experience any other delays under these alternative strategies. The

[8] For interested readers, Petit et al. (2018) studied scenarios where the dynamic substitution strategy uses the same number of buses in the fleet as the one- and two-way looking strategies, but with various splits between operating buses and standby buses.

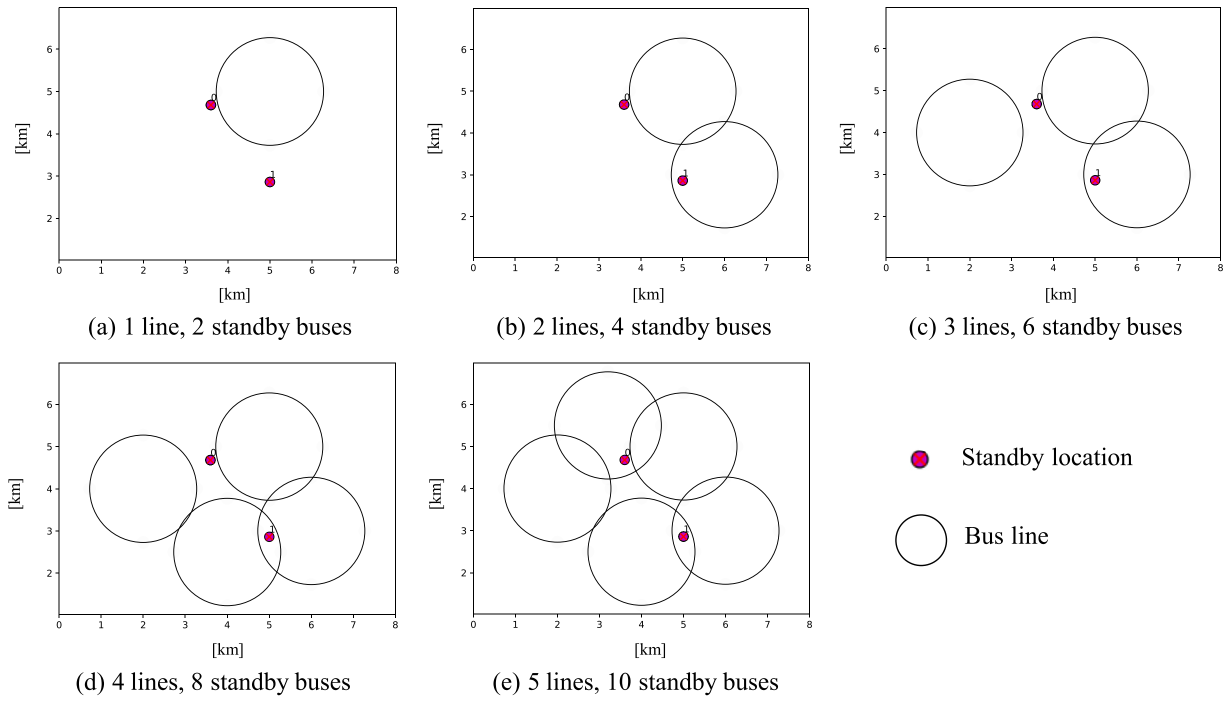

Fig. 3. System layout of the hypothetical examples.

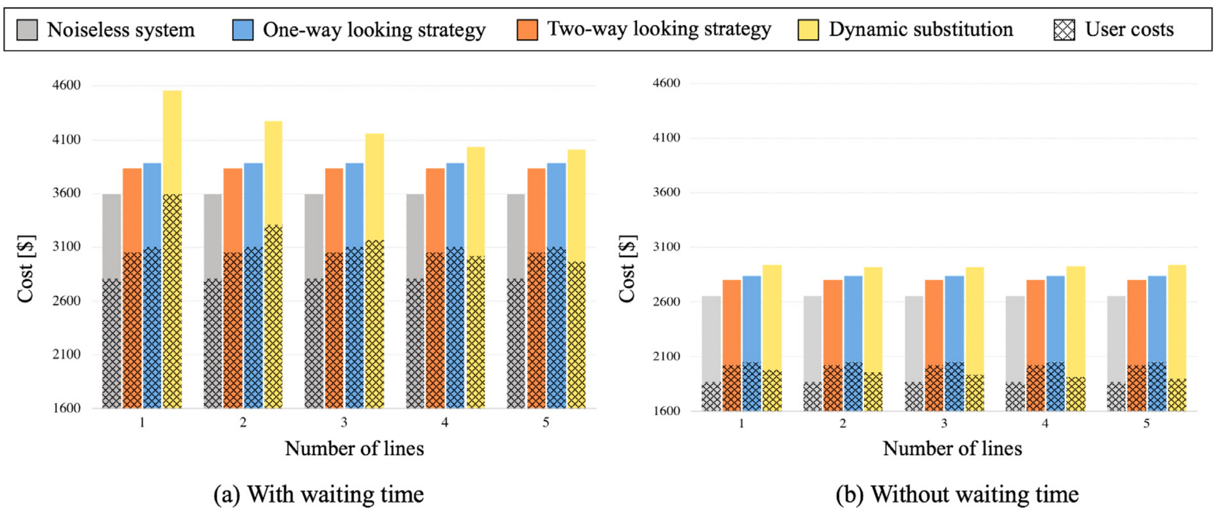

Fig. 4. Costs per line for the different control strategies.

waiting and riding costs for the dynamic substitution strategy are obtained through simulation of cumulative boardings and alightings of each operating and NIS bus. All costs are averaged over a set of 20 simulations. As mentioned in Petit et al. (2018), the substitution strategy can compete with the most advanced control strategies (e.g., speed control), especially in the case of driverless vehicles. Therefore, here we assume that all buses are driverless vehicles and the agency costs only include the costs of operating the vehicles, along with the depreciation costs. The cost of self-driving technology for those autonomous vehicles is ignored. In addition, we consider the alternative case where passengers have access to real-time bus arrival times. Thereby, their waiting time can be reduced to a negligible amount.

Fig. 4(a) shows that the overall system costs per bus line decrease as the number of lines in the system increases, which indicates that the dynamic substitution strategy can benefit from the "economies of scale" effect for pooling the standby fleet across the entire system. On the other hand, the speed control strategies are implemented on each line independently of each other and do not benefit from such economies; their unit costs remain constant across all scenarios. One can also

Table 1
Average performance for different network sizes and different control methods (4-hr time period).

	Number of lines	1	2	3	4	5
Noiseless system	Waiting costs	937	1873	2810	3746	4683
	Riding costs	1873	3746	5620	7493	9366
	Total passenger costs	2810	5620	8429	11239	14049
	Agency costs	784	1568	2352	3136	3920
	Total costs	3594	7188	10781	14375	17969
	Total costs per line			3594		
One-way looking strategy (Daganzo, 2009)	Waiting costs	1050	2100	3150	4199	5249
	Riding costs	2053	4107	6160	8214	10267
	Total passenger costs	3103	6207	9310	12413	15517
	Agency costs	784	1568	2352	3136	3920
	Total costs	3887	7775	11662	15549	19437
	Total costs per line			3887		
Two-way looking strategy (Daganzo and Pilachowski, 2011)	Waiting costs	1028	2057	3085	4114	5142
	Riding costs	2021	4042	6063	8084	10105
	Total passenger costs	3050	6099	9149	12198	15248
	Agency costs	784	1568	2352	3136	3920
	Total costs	3834	7667	11501	15334	19168
	Total costs per line			3834		
Dynamic substitution	Waiting costs	1612	2715	3708	4427	5354
	Riding costs	1982	3913	5817	7657	9496
	Total passenger costs	3594	6628	9525	12084	14850
	Agency costs	880	1760	2640	3520	4400
	Substitution costs	79	166	319	553	810
	Total costs	4553	8554	12484	16157	20060
	Total costs per line	4553	4277	4161	4039	4012

see that when $|\mathcal{K}| \geq 4$, the proposed substitution strategy outperforms both speed control strategies in terms of passenger costs. As a result, although it requires additional resources from the agency, its overall system costs do not exceed those of the best speed control strategies by more than 5% when $|\mathcal{K}| \geq 4$. The detailed numerical results in Table 1 indicate that the system costs per line decrease by 6.1% for the two-line system, 8.7% for the three-line system, 11.3% for the four-line system, and 11.9% for the five-line system. This clearly indicates that for a massive transit system with many lines, the proposed substitution strategy holds the greater potential to outperform the most advanced control methods to date by sharing a fleet of standby buses across the system. It is also interesting to see that in the five-line case, the passenger costs for the substitution strategy are within 6% of those of the noiseless system; the standby resources almost balance out all the inconvenience caused by the disturbances.

In Fig. 4(b), waiting costs are ignored due to availability of real-time information. Although the benefit from the pooling effect is reduced, the substitution strategy benefits the most. As a result, it outperforms both speed control strategies in terms of passenger costs for all cases. Note that the performance of both speed strategies is slightly better too, as we ignore the extra waiting due to slower commercial speeds.

To gain more insights on the utilization of the standby buses, it is also interesting to note that their average usage during the 4-hour period increases as the number of lines/buses increases, from about 25% for one line to nearly 55% for five lines. This is because the agency has more standby buses at hand, and can better react to schedule deviations. When standby buses are left unused for a significant amount of time, as shown in this example, it opens the opportunity for re-using the standby buses as regular buses from time to time. The passenger experience would be further improved, while the agency makes better use of the expensive standby resources. While a dynamic optimization of the standby fleet might be difficult to implement in short term windows, it would still be beneficial for the agency to adapt its fleet configuration during a longer time horizon. This is the focus of the next example.

5.2. Dynamic fleet management

When a significant amount of time exists for standby buses to remain idle, we may improve their usage by reducing the fleet size as well as enhancing their utilization. This is especially the case when systems face time-varying demand over a multi-period horizon.

We consider two homogeneous lines with length $L_k = 9$ km over two 4-hour periods with two standby locations, similar to Fig. 3(b). The first time period represents an off-peak period (e.g., 1 pm - 5 pm) with demand $\lambda_k = 30$ pax/km-h, and the second represents a peak period (e.g., 5 pm - 9 pm) with $\lambda_k = 50$ pax/km-h. As travel demand typically alternates between peak and off-peak periods, the two time periods can be representative of a typical “peak/off-peak” cycle within a day of service. We assume that the agency wants to ensure a minimum scheduled headway of 5 min during peak hours, and 8 min during off-peak. Therefore, there should be at least 3 and 5 buses operating on each line during off-peak and during peak-hours, respectively. It is assumed that the agency can have at most 18 buses available to serve those two lines and each bus would incur a prorated acquisition cost of 16 $/h. The remaining parameters are the same as in Section 5.1. As

Table 2
System costs and optimal fleet configurations for different fleet sizes.

Total fleet size	Off-peak period				Peak period			Total cost
	Operating fleet	Standby fleet	Unused	Costs	Operating fleet	Standby fleet	Costs	
12	(4,4)	4	0	8,639	(5,5)	2	Inf	Inf
13	(4,4)	5	0	8,518	(5,5)	3	13,819	22,337
14	(4,4)	5	1	8,606	(5,5)	4	13,300	21,906
15	(4,4)	5	2	8,694	(5,5)	5	13,355	22,049
16	(4,4)	5	3	8,782	(5,5)	6	13,263	22,045
17	(4,4)	5	4	8,870	(6,6)	5	13,172	22,042
18	(4,4)	5	5	8,957	(6,6)	6	13,142	22,099

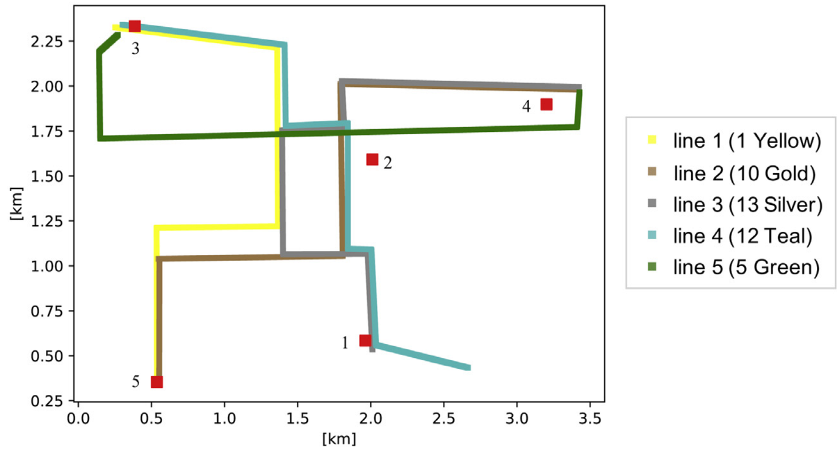

Fig. 5. CUMTD 5-line network.

Table 3
Characteristics of the five CUMTD lines.

Line	1	2	3	4	5
Length [km]	7.52	9.01	8.22	8.06	8.20
Number of buses	3	3	3	3	3
Demand density [pax/km-hr]	35.0	30.0	50.0	32.0	40.0
Headway [min]	6.8	7.7	8.8	7.0	8.0

such, all operating and standby buses are associated with a depreciation cost of 12$/h. An unused bus would be associated with a depreciation cost of 6 $/h, i.e., 50% of the operating bus depreciation value.

Table 2 shows the optimal fleet configuration for each fleet size between 12 and 18 buses. The costs reported include the waiting costs, riding costs, substitution costs and fixed agency costs to account for operations, depreciation, and acquisition. Note that for a fleet of 12 buses during the peak period, two standby buses are not sufficient to prevent the system from failing; i.e., buses end up bunching permanently and substitutions cannot occur frequently enough to restore the system. It is also interesting to see how during the off-peak period, it is preferable to leave buses unused rather than acquiring them in the standby fleet. Their lower depreciation costs outrun the marginal benefit that they could provide if they were in the standby fleet. The optimal fleet size appears to be 14 buses, with 5 operating buses dispatched on each line and 4 left as standby during the peak period, and 4 operating buses on each line, with 5 standby buses and one unused during the off-peak period. It is noticeable that the agency should transform one standby bus along with the unused one in the off-peak period into operating buses in the peak period. In this way, the opportunity costs for maintaining a standby fleet are reduced through effective dynamic fleet management across multiple periods.

5.3. CUMTD case study

In the following, we consider a real-world test case with five of the most popular CUMTD bus lines on the campus of the University of Illinois at Urbana-Champaign (UIUC); i.e., 1 Yellow, 10 Gold, 13 Silver, 12 Teal and 5 Green. The topological layout of the bus routes is shown in Fig. 5, which is drawn to scale. The parameters of each line are displayed in Table 3. All distances between points are computed with the L_1 norm to account for the fact that streets form a dense grid

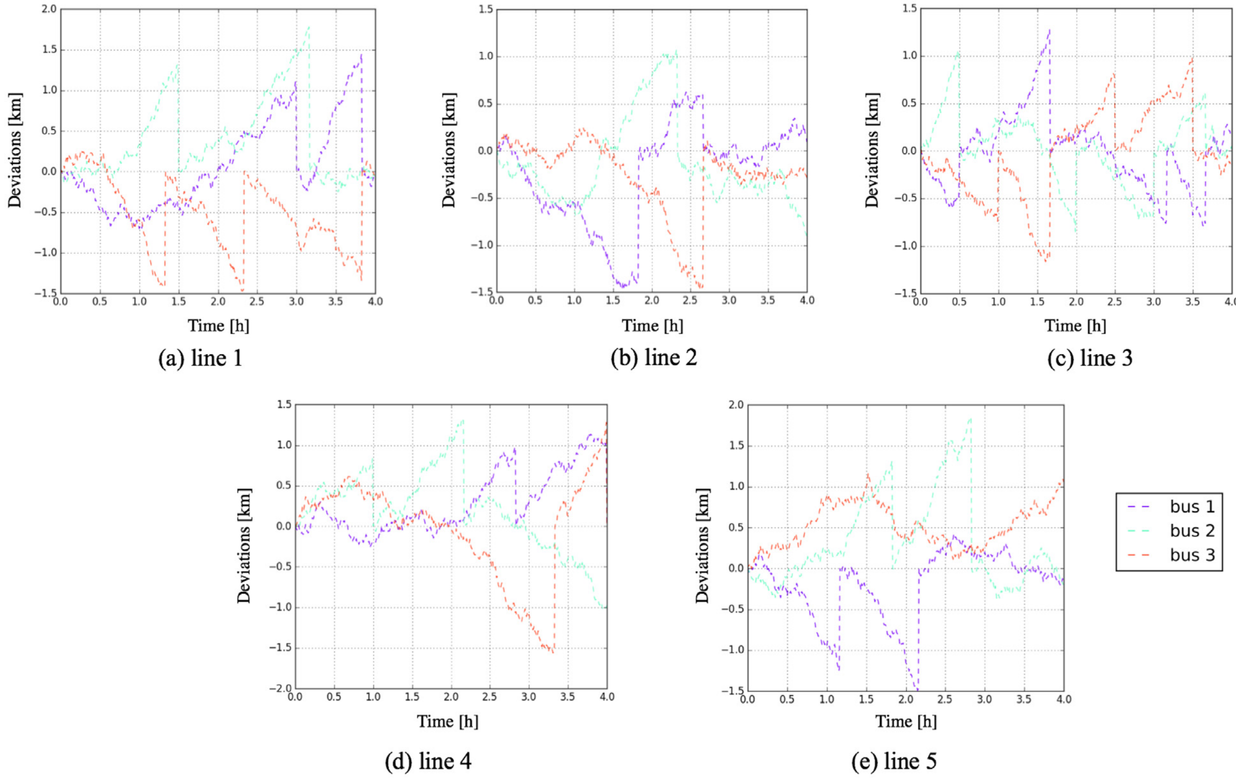

Fig. 6. Evolution of schedule deviations under substitutions.

on the campus.[9] Five large parking lots are selected as candidate locations for holding standby buses, as marked by red squares in Fig. 5. The cruising speed for all buses is set to be $E = 25$ km/h by taking account of a large number of stops on campus and high density of interfering pedestrians. Boarding and alighting times are assumed to be $B = 4$ s/pax and $A = 2$ s/pax, respectively. The monetary value of passenger time is assumed to be $\mu = 10$ \$/h (i.e., about 50% of the hourly wage in Champaign-Urbana (Bureau of Labor Statistics, 2017)) and the operating cost factor is assumed to be $c_M = 37$ \$/h. We consider using a fleet of 6 standby buses to control bunching across these five lines in one time period.

Fig. 6 illustrates the deviations from the schedule for all five lines. It shows that the deviations can be contained within a reasonable limit by using the six standby buses. The maximum deviation on lines 2 to 4 is less than 1.5 km, which corresponds to about 3.5 min of lateness/earliness. For lines 1 and 5, the deviations are also usually contained below 1.5 km but can occasionally reach 1.8 km (i.e., less than 4 min). Fig. 7 illustrates the occupancy of the operating buses on line 3, the highest across all lines. We can observe that it reaches a maximum of about 27 passengers, which can be handled easily by a regular single-deck bus used by CUMTD.

To obtain an overview of the usage and distribution of standby buses, we keep track of the number of standby buses at each parking location over time, as shown in Fig. 8. First, we clearly see that on average half of the standby fleet is in NIS mode at all times. This gives a lower bound of the number of standby buses needed to operate this 5-line system. We can also observe that the maximum average number of standby buses at a given location at the same time is about 2 (at location 2), which can easily be accommodated by most of the parking lots on UIUC campus. On average, there are 14.0 substitution dispatches from location 2 during the 4-hr planning horizon. Location 2 appears as a strategic location as it is the most frequently used place for holding standby buses. This is intuitive since it is the only centrally located one that can offer the widest coverage and the fastest substitution to all five lines. In comparison, locations 1, 3, 4, and 5 respectively yield 3.5, 4.4, 2.9, and 2.5 dispatches on average. Locations 1, 4 and 5 are used least frequently because of their peripheral positions with respect to the overall layout of the transit lines. Location 3 has the advantage of being close to 3 lines, which explains why it is used slightly more often than locations 1, 4 and 5. Therefore, we can draw the insights that, to fully exploit the benefit of the substitution strategy, a transit agency should ensure proper coverage of the system by reserving certain parking spots near the places where most of the bus lines would intersect.

[9] One-way street restrictions and other geographical obstacles are ignored for simplicity.

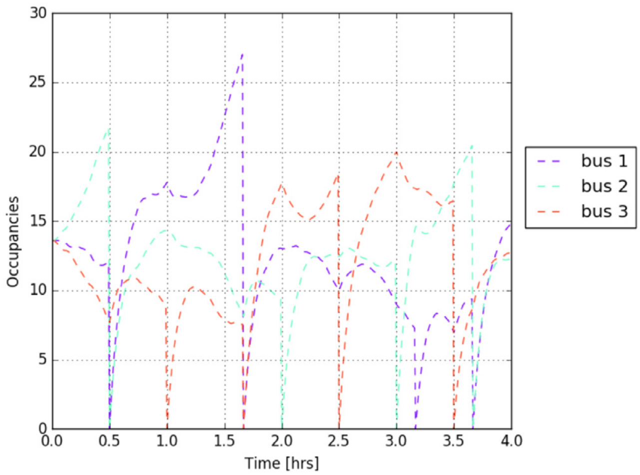

Fig. 7. Evolution of bus occupancies with substitution on line 3.

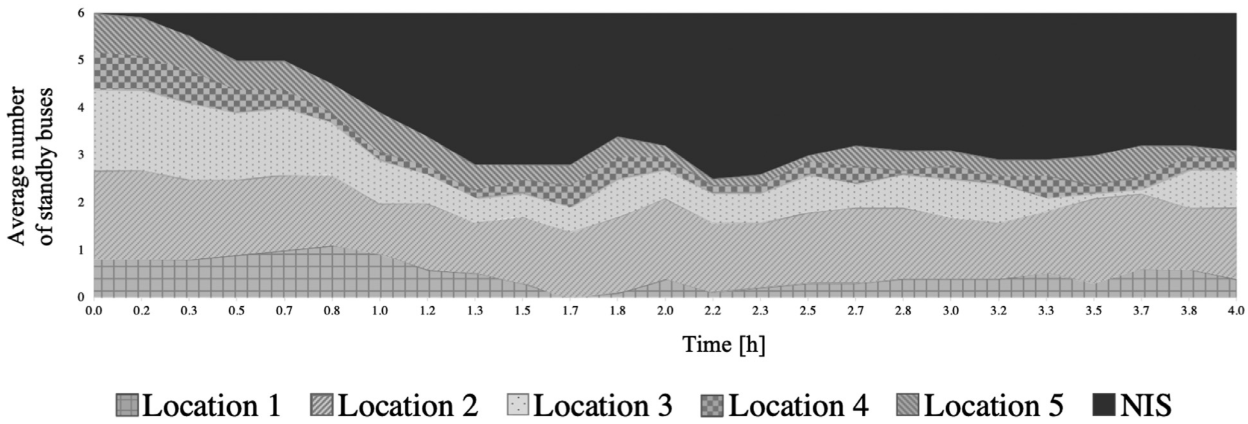

Fig. 8. Distribution of standby locations over the 4-hour time horizon.

5.4. Hybrid bunching control

We know that the dynamic holding/speed strategies in the literature can effectively absorb small schedule deviations up to a certain magnitude. However, adverse weather conditions, persistent traffic congestion, or surge in travel demand, can cause large disturbances that push the system out of the stable domain of these strategies. When such perturbations happen, those strategies will likely fail to keep the transit system on time. The dynamic substitution strategy can serve as a perfect complement to the dynamic holding/speed strategies, by triggering a substitution as a remedy once the deviations become so large that either the system will need a significant amount of time to recover, or it will simply fail. Meanwhile, the dynamic holding/speed control strategies can also help smooth out small perturbations such that substitutions will be used only infrequently — and only a smaller number of standby buses will be needed.

In the following, we combine the bus substitution with the one-way looking strategy proposed in Daganzo (2009) as an example of the hybrid method, and implement it on the CUMTD case as described in Section 5.3. In addition, we assume that buses cannot be early with respect to their schedules, to help maintain schedule stability.

To represent a heavy load, the demand densities for the five lines are set to be 79, 75, 90, 77, and 83 pax/km-hr, respectively. A larger disturbance and a greater value of time are employed as well, where $\sigma_k = 0.147$ km/min and $\mu = 30$ \$/h, respectively. This could represent a busy weekday where many students commuting to class cannot be too late. Since speed control is implemented, we only consider a fleet of three standby buses. The remaining parameters are the same as in section 5.3.

According to Daganzo and Pilachowski (2011), the dynamics of the bus trajectories are

$$y_{k,n}(t+1) = y_{k,n}(t) + V_k - \delta_k + \alpha_k s_{k,n}(t),\ \forall t \in \{0, 1, 2 \cdots\}, k \in \mathcal{K}, n \in \mathcal{N}_k,$$

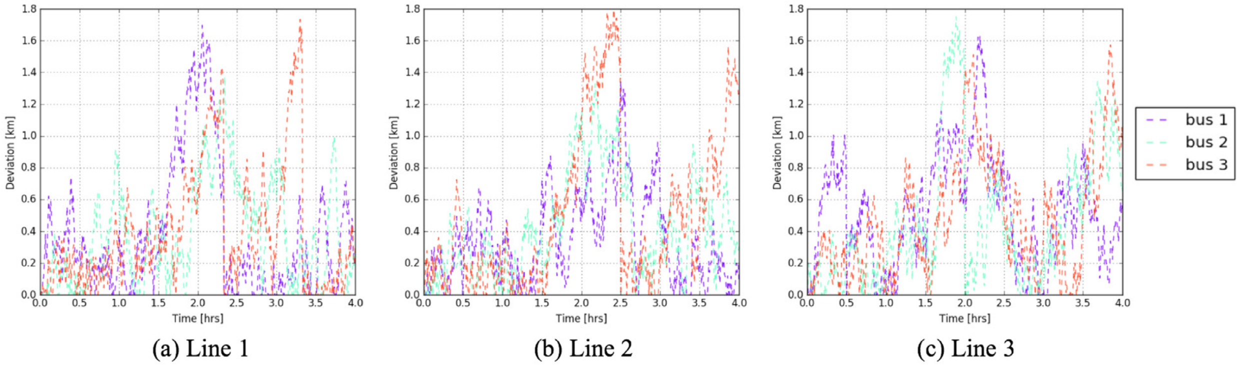

Fig. 9. Sample trajectories of 3 lines under speed control and substitutions.

Table 4
Average results for different standby fleet sizes and different control methods (4-hr time horizon).

	Average waiting cost	Average riding cost	Total passenger costs	Operating costs		Substitution costs	Total costs	
				w/ driver	w/o driver		w/ driver	w/o driver
One-way looking strategy	8,765	43,946	52,711	6,720	2,940	0	59,431	55,651
Hybrid strategy	7,569	43,090	50,659	7,620	3,084	150	58,429	53,893

where α_k and δ_k are the control parameters. Passengers check the schedule and arrive at the time that the bus is scheduled to arrive. As such, the passenger waiting time is simply proportional to the schedule deviations. The formula for the number of passengers boarding bus n on line k at time t is $\lambda_k \cdot \frac{S_k}{V_k} \cdot v_{k,n}(t)$ and their associated waiting time is $\lambda_k \frac{S_k}{V_k} \epsilon_{k,n}(t)$.

A set of sample trajectories are shown in Fig 9. One can observe that deviations are effectively contained by only a few substitutions despite the high demand and large disturbances. Most substitutions occur when the deviations reach about 1.3 km (e.g., line 1 at t =3.3 hours, line 2 at t = 2.5 hours), or when buses are not evenly spaced (e.g., line 3 at t = 0.5 hours).

Table 4 reports that the passenger costs of the proposed hybrid approach decrease by nearly 4% as compared to that of the speed control strategy alone. If we assume that drivers operate the buses at a cost of 63 $/h, the agency's costs for applying the hybrid method would increase by 16%. However, due to the huge demand and the large value of time, the overall system costs can be reduced by 1.7%. If we consider driverless vehicles, system costs are reduced by a higher value of 3.2%. Thus, integration of the substitution strategy with other mitigation approaches shows a great promise to maintain a more reliable service for the transit systems that are heavily loaded or suffer from serious disturbances.

5.5. Impact of finite vehicle capacity

As mentioned in Section 2.1, this paper assumes that buses have infinite capacity, and hence vehicle substitutions will not be induced by full buses. Such a simplification will be reasonable only if the schedule deviation and bus occupancy are highly correlated (e.g., late buses tend to carry more passengers, and vice versa). To test whether such an assumption is reasonable, we now impose a finite bus capacity which may also trigger vehicle substitutions, and investigate its impact on the system performance.

To ensure a high occupancy of buses, we consider two identical bus lines, each with length $L = 9$ km, $N_k = 3$ operating buses, and a demand density of $\lambda = 80$ pax/km-h. These two lines share 4 standby buses. The remaining settings are the same as the two-line case in Section 5.1. We apply the methodology described in Section 3 to obtain the VFA policy (i.e., assuming infinite vehicle capacity). We then simulate the system operations with this VFA policy, but now consider a finite vehicle capacity that ranges from 50 to 75 passengers. Passengers are denied boarding once a bus is full, and they have to wait for the next available bus. At each decision epoch, if an operating bus has reached its capacity, a substitution is triggered with the closest available standby bus. The simulations are run over 50 realizations of uncertainties; for each vehicle capacity, the average number of substitutions and the corresponding average system costs are reported in Table 5. The substitutions are further classified into 3 categories: (i) those only due to deviations (based on the VFA policy), although this operating bus has not reached its capacity; (ii) those only due to occupancy, i.e., a bus has reached its capacity but the VFA policy does not trigger a substitution; (iii) both, i.e., the VFA policy triggers a substitution and meanwhile this bus has reached its capacity. The corresponding performance under infinite vehicle capacity is also included as the benchmark.

Table 5 shows that, in general, more substitutions are triggered by both the VFA and the capacity constraint as the bus capacity decreases. This shows a strong correlation between schedule deviation and bus occupancy. As a result, the number of substitutions due to occupancy only is very small as compared to those covered by the VFA policy. Also, the total number of substitutions only increases marginally (i.e., by 5%) for lower-capacity vehicles, which indicates that substitutions due to

Table 5
Average results for different vehicle capacities (4-hr time horizon).

Capacity	Substitutions due to			Total substitutions	Waiting costs	Riding costs	Substitution costs	Total costs
	Deviations	Occupancy	Both					
Inf	22.7	0.0	0.0	22.7	14,508	13,012	278	27,798
75	22.1	0.0	0.7	22.8	14,483	12,973	276	27,732
70	21.8	0.0	1.0	22.8	14,530	12,971	276	27,777
65	21.2	0.0	1.6	22.8	14,602	12,968	276	27,846
60	20.3	0.1	2.7	23.1	14,674	12,954	276	27,904
55	18.6	0.4	4.4	23.4	14,854	12,941	278	28,073
50	16.9	0.9	6.2	24.0	15,088	12,902	279	28,269

occupancy are "dominated" by VFA-triggered substitutions for these cases. Not surprisingly, the average total costs increase by less than 2% when the vehicle capacity drops from infinity to 50.

Notice too, interestingly, that the total costs for cases with certain finite vehicle capacities (70-75) are slightly smaller than those of infinite capacity. In those cases, the additional wait that passengers experience (due to full buses) is outweighed by the shorter turn-around time that an NIS bus may need so as to become available again for the next substitution (since there is likely a smaller number of onboard passengers that must alight before the bus can retire).

One should recognize that the above conclusion only holds for the chosen system parameters. In practice, a transit agency may face very low vehicle capacities that impact vehicle substitution more significantly. In such cases, our model can still be applicable if we still include the vehicle occupancy as part of the system state and Bellman Equation (e.g., (23)), and train the VFA policy accordingly.

6. Conclusion

This paper uses a dynamic substitution strategy to mitigate bus bunching for urban transit systems with multiple bus lines. This strategy entails dispatching standby buses, which are prepositioned at suitable locations, to take over service from late/early in-operation buses. Once an operating bus has been substituted, it changes into NIS or retiring mode, in which only passenger alighting is allowed until the bus becomes empty. Therefore, the agency needs to decide how to dispatch the shared pool of standby buses to substitute operating buses across all lines, and where to reposition the NIS buses over the service region. The agency's goal is to deliver the best service, e.g., least waiting and dwelling times of passengers, while minimizing the total operating costs. Under homogeneous settings, we formulate the problem into a stochastic dynamic program in an infinite-horizon and develop a non-myopic ADP-based algorithm to reveal the optimal decision-making policy for substitution and repositioning. We also present a multi-period dynamic program to tackle time-varying systems with a dynamic utilization of the fleet (either to provide regular service or serve as standby). The proposed strategies and solution algorithms are tested with a set of hypothetical transit networks, and compared with the existing adaptive speed control methods under homogeneous settings. Results show that the dynamic substitution strategy can benefit from "economies of scale" by pooling standby buses across multiple lines, and outperform its counterparts in terms of decreasing passenger costs. We also quantify the benefit of dynamic fleet management under time-varying demand. It is shown that optimal fleet utilization allows further savings on vehicle opportunity costs. Moreover, a case study based on a subnetwork of the transit system in Champaign, Illinois, is also presented, showing the potential of the proposed approach for stabilizing small transit systems. Finally, a hybrid approach that combines speed control and bus substitution strategies is introduced to stabilize extremely unstable systems. It is shown that it yields smaller system costs than any single strategy used alone.

In this research, the proposed strategy is tested in small-scale networks. However, in reality transit agencies may monitor over 50 or more bus lines. It shall be intuitive that the larger the number of bus lines, the higher the benefits from pooling standby buses. Therefore, one of the future research opportunities could be to apply the proposed dynamic substitution strategy to large-scale real-world transit networks. One challenge in this direction has to do with solving the Bellman equation for many lines, as the optimization problem at each decision epoch involves a difficult combinatorial problem when the transit network becomes large. Future work should also seek field implementations of the proposed strategy, to examine the real performance of the strategy as compared to the theoretical predictions. Also, future efforts can be devoted to improving the proposed hybrid approach under large disturbances. The design of the speed control law could account for the impact of the bus substitutions, especially when deviations grow large. This would limit the loss in cruising speed and improve the passenger experience. Furthermore, vehicle capacity constraints could be considered in the model to also allow for substitutions when vehicle occupancy is reaching the maximum capacity. In that case, vehicle occupancy should also be included in the VFA as part of the system state. The proposed dynamic substitution strategy could also potentially be modified to control headway-based systems; e.g., by using headways as state variables and determining bus insertion locations/times based on bus trajectory projections. Finally, we are also interested in the strategic planning of the standby fleet across a network of routes. We wish to optimize the allocation and positioning of standby resources in a large-scale transit network to reduce the agency costs while maintaining a satisfying level of service. In addition, we could consider the transit demand as endogenously dependent on the level of service; e.g., the total amount of transit riders is modeled as a function of passenger experience (i.e., travel time), and included as part of the state variables that influence the system dynamics.

Acknowledgments

The authors thank Mr. Jay Rank and his colleagues at CUMTD for providing helpful information on their business operations. The work on dynamic fleet management across multiple periods (Section 4) was partially motivated by a comment from one of the anonymous reviewers of Petit et al. (2018) on opportunity costs of the standby buses. We also thank the two anonymous reviewers of this current paper for their helpful comments. This work was financially supported in part by the U.S. National Science Foundation through Grant CMMI-1662825, as well as a grant from the U.S. Army Corps of Engineers.

Appendix A. Status of the not-in-service buses

Recall $u^R_{k,n,j}(t) \in \{0,1\}, \forall k \in \mathcal{K}, n \in \mathcal{N}_k, j \in \mathcal{J}$ denote whether or not NIS bus (k,n) reaches parking location j before the next decision epoch, i.e., $u^R_{k,n,j}(t) = 1$ if NIS bus (k,n) is going to reach parking location j before $t + \Delta$, and $u^R_{k,n,j}(t) = 0$ otherwise. The relationship between $u^R_{k,n,j}(t)$ and $\tau^R_{k,n,j}(t)$ can be expressed by the following set of inequalities:

$$u^R_{k,n,j}(t) \geq \alpha^1_{k,n,j}(t) + \alpha^2_{k,n,j}(t) - 1,\ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\}, \tag{27}$$

$$u^R_{k,n,j}(t) \leq \alpha^1_{k,n,j}(t),\ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\}, \tag{28}$$

$$u^R_{k,n,j}(t) \leq \alpha^2_{k,n,j}(t),\ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\}, \tag{29}$$

where $\alpha^1_{k,n,j}(t)$ and $\alpha^2_{k,n,j}(t)$ are binary indicators defined as follows:

$$\alpha^1_{k,n,j}(t) \geq \frac{\tau^R_{k,n,j}(t)}{M},\ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\}, \tag{30}$$

$$\alpha^1_{k,n,j}(t) < 1 + \frac{\tau^R_{k,n,j}(t)}{M},\ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\}, \tag{31}$$

$$\alpha^2_{k,n,j}(t) > \frac{\Delta - \tau^R_{k,n,j}(t)}{M},\ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\}, \tag{32}$$

$$\alpha^2_{k,n,j}(t) \leq 1 + \frac{\Delta - \tau^R_{k,n,j}(t)}{M},\ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\}, \tag{33}$$

$$\alpha^1_{k,n,j}(t), \alpha^2_{k,n,j}(t) \in \{0,1\},\ \forall j \in \mathcal{J}, k \in \mathcal{K}, n \in \mathcal{N}_k, t \in \{0,1,2,\dots\}, \tag{34}$$

where M is an upper bound of $\tau^R_{k,n,j}(t), \forall t, k \in \mathcal{K}, n \in \mathcal{N}_k, j \in \mathcal{J}$. Eqs. (30)-(34) state that $\alpha^1_{k,n,j}(t) = 1$ if and only if $\tau^R_{k,n,j}(t) > 0$, and $\alpha^2_{k,n,j}(t) = 1$ if and only if $\tau^R_{k,n,j}(t) \leq \Delta$, such that $u^R_{k,n,j}(t) = 1$ if and only if $\tau^R_{k,n,j}(t) \in (0, \Delta]$.

Appendix B. Continuous vs. discrete models

In this paper, we model the spatial dimension with a continuous approximation. We now compare this continuous model with a more commonly seen discrete model. To be consistent with the rest of the paper, we denote the demand density in the discrete system by λ and the commercial speed of the buses by V. In addition, we denote the stop spacing by s, and the walking speed by v_w.

First, we look at a discrete system with no randomness, where the target headway H is maintained. Under such conditions and with discrete stops, the total waiting time, W_1^{D}, experienced by the passengers within the catchment area illustrated in Fig. 10, can be expressed as

$$W_1^{\mathrm{D}} = \int_{x=0}^{s/2} \left[\int_{t=x/V}^{x/V+H} \left(H - t - \frac{x}{V} \right) \lambda dt + \int_{t=H-x/v_w}^{x/V+H} \left(2H - t - \frac{x}{V} \right) \lambda dt \right] dx$$
$$+ \int_{x=-s/2}^{0} \left[\int_{t=x/V}^{x/v_w+H} \left(H - t + \frac{x}{V} \right) \lambda dt + \int_{t=H+x/v_w}^{x/V+H} \left(2H - t + \frac{x}{V} \right) \lambda dt \right] dx, \tag{35}$$

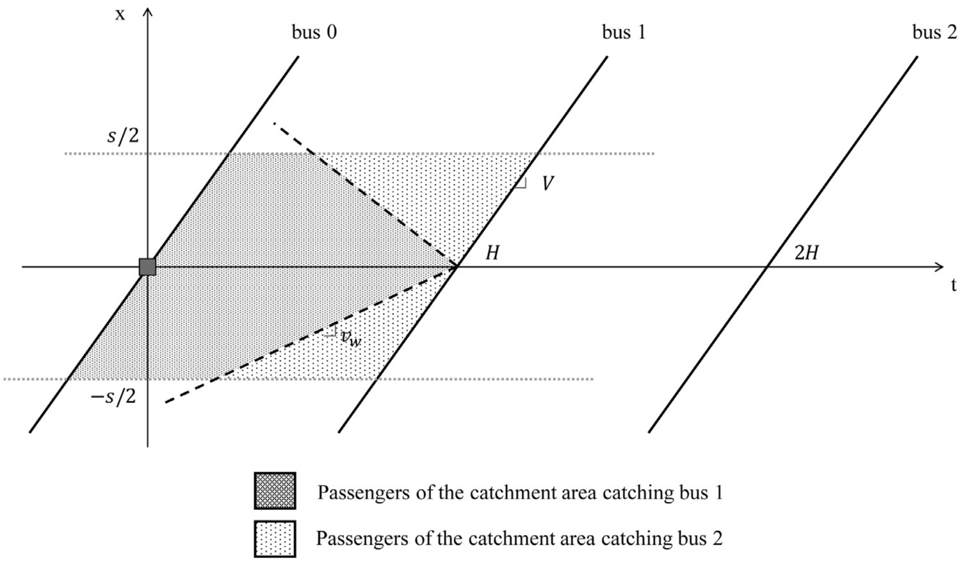

Fig. 10. Catchment area over one stop spacing.

where the first term corresponds to the passengers within this area who are able to catch bus 1, and the second term pertains to the passengers who have to wait for bus 2. Simple algebra shows that the above formula reduces to the following:

$$W_1^{\mathrm{D}} = \frac{\lambda H^2 s}{2}. \tag{36}$$

In the continuous model, the equivalent demand in this catchment area is cleared when a bus travels a distance of s. At each time step, a bus travels a distance of V (in units of distance per time step). As a result, the number of passengers boarding within this time step is λVH, the average waiting time is $H/2$, and the number of time steps to travel a distance of s is about s/V. As such, the total waiting time experienced by passengers within the same catchment area in the continuous model, denoted by W_1^{C}, equals to $\frac{\lambda H^2 s}{2}$, which is exactly the same as W_1^{D}. Moreover, in the discrete model, the travel time for each passenger can be approximated by

$$T_1^{\mathrm{D}} \approx \frac{L}{4V} + \frac{s}{2v_w} - \frac{s}{2V}. \tag{37}$$

If $L \gg s$, then Eq. (37) yields the same result as that of the continuous model, i.e., $\frac{L}{4V}$.

Now let's consider the case when randomness arises. For the discrete model, the extra waiting time W_2^{D} and extra travel time T_2^{D} experienced by an average passenger, due to randomness, have been estimated in Daganzo and Ouyang (2019) as follows:

$$W_2^{\mathrm{D}} = \frac{1}{2} \cdot \frac{Var(h)}{H},$$
$$T_2^{\mathrm{D}} \approx \left[(1+\beta)^S - 1\right] \cdot \frac{Var(h)}{H}, \tag{38}$$

where S is the average number of stops that a passenger travels, h is the actual headway that passengers experience in the vehicle, and $\beta \ll 1$ is a unitless constant that can be approximated as λsB (where B is the time needed for boarding one passenger). A similar analysis on the continuous model shows that the average extra waiting time per passenger $W_2^{\mathrm{C}} = W_2^{\mathrm{D}}$, and that

$$T_2^{\mathrm{C}} \approx \left[\left(1+\frac{\beta}{k}\right)^{kS} - 1\right] \cdot \frac{Var(h)}{H}, \tag{39}$$

where k is the number of time steps needed to overcome one stop spacing ($k \gg 1$). The decomposition of T_2^{D} and T_2^{C} down to the second order, using the Binomial theorem, yields

$$T_2^{\mathrm{D}} \approx \left(S\beta + \frac{S(S-1)}{2}\beta^2\right) \cdot \frac{Var(h)}{H},$$
$$T_2^{\mathrm{C}} \approx \left(S\beta + \frac{S(kS-1)}{2k}\beta^2\right) \cdot \frac{Var(h)}{H}. \tag{40}$$

This clearly shows that for $\beta \ll 1$, the two terms are approximately equal. As $k \to \infty$, the difference in the second order term is in order of $\frac{1}{S}$, which is about 5% to 10%, for typical values of $S \sim 10-20$ stops. As a result, the estimated travel time in the continuous model is slightly larger than that in the discrete model.

In addition, Pilachowski (2009) has used simulations to compare the bunching phenomena in both a discrete-stop setting and a continuous-stop setting. It was found that when the system is left uncontrolled, the buses in the continuous stop model collapse into bunches earlier than the discrete stop model. This indicates that a system with discrete stops is easier to control than the continuous counterpart, and that less substitutions would be needed, which also reduces agency costs for substitutions. In light of those conclusions, the proposed strategy can be expected to perform better in discrete settings.

References

Abkowitz, M.D., Lepofsky, M., 1990. Implementing headway-based reliability control on transit routes. Journal of Transportation Engineering 116 (1), 49–63.

Argote-Cabanero, J., Daganzo, C.F., Lynn, J.W., 2015. Dynamic control of complex transit systems. Transportation Research Part B: Methodological 81 (1), 146–160.

Bartholdi, J.J., Eisenstein, D.D., 2012. A self-coördinating bus route to resist bus bunching. Transportation Research Part B: Methodological 46 (4), 481–491.

Berrebi, S.J., Watkins, K.E., Laval, J.A., 2015. A real-time bus dispatching policy to minimize passenger wait on a high frequency route. Transportation Research Part B: Methodological 81, 377–389.

Bureau of Labor Statistics, 2017. May 2016 metropolitan and nonmetropolitan area occupational employment and wage estimates Champaign-Urbana, IL. https://www.bls.gov/oes/current/oes_16580.htm.

Ceder, A., 1990. Optimal design of transit short-turn trips. Transportation Research Record: Journal of the Transportation Research Board 1221, 8–22.

Daganzo, C., Ouyang, Y., 2019. Public Transportation Systems Public Transportation Systems: Principles of System Design, Operations Planning and Real-Time Control. World Scientific.

Daganzo, C.F., 2008. Fundamentals of Transportation and Traffic Operations. Emerald, Inc..

Daganzo, C.F., 2009. A headway-based approach to eliminate bus bunching: systematic analysis and comparisons. Transportation Research Part B: Methodological 43 (10), 913–921.

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

Delgado, F., Munoz, J.C., Giesen, R., 2012. How much can holding and/or limiting boarding improve transit performance? Transportation Research Part B: Methodological 46 (9), 1202–1217.

Delgado, F., Muñoz, J.C., Giesen, R., Cipriano, A., 2009. Real-time control of buses in a transit corridor based on vehicle holding and boarding limits. Transportation Research Record: Journal of the Transportation Research Board (2090) 59–67.

Eberlein, X.J., Wilson, N.H.M., Bernstein, D., 2001. The holding problem with real-time information available.. Transportation Science 35 (1), 1–18.

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

Fu, L., Liu, Q., Calamai, P., 2003. Real-time optimization model for dynamic scheduling of transit operations. Transportation Research Record: Journal of the Transportation Research Board (1857) 48–55.

Hernández, D., Muñoz, J.C., Giesen, R., Delgado, F., 2015. Analysis of real-time control strategies in a corridor with multiple bus services. Transportation Research Part B: Methodological 78, 83–105.

Ibarra-Rojas, O., Delgado, F., Giesen, R., Muñoz, J., 2015. Planning, operation, and control of bus transport systems: A literature review. Transportation Research Part B: Methodological 77, 38–75.

Knuth, D., 1972. The Art of Computer Programming. Addison-Wesley.

Ling, K., Shalaby, A., 2004. Automated transit headway control via adaptive signal priority. Journal of Advanced Transportation 38 (1), 45–67.

Liu, H., Skabardonis, A., Zhang, W., 2003. A dynamic model for adaptive bus signal priority. In: Preprint CD-ROM, 82nd Transportation Research Board Annual Meeting, Washington, DC.

Liu, Z., Yan, Y., Qu, X., Zhang, Y., 2013. Bus stop-skipping scheme with random travel time. Transportation Research Part C: Emerging Technologies 35, 46–56.

Neff, J., Dickens, M., 2015. 2013 Public Transportation Association Fact Book. Technical Report. American Public Transportation Association, Washington, DC.

Newell, G.F., 1974. Control of pairing of vehicles on a public transportation route, two vehicles, one control point.. Transportation Science 8 (3), 248–264.

Newell, G.F., Potts, R.B., 1964. Maintaining a bus schedule. Proceedings of the 2nd Australian Road Research Board Conference 2, 388–393.

Osuna, E.E., Newell, G.F., 1972. Control strategies for an idealized public transportation system.. Transportation Science 6 (1), 52–72.

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

Pilachowski, J.M., 2009. An Approach to Reducing Bus Bunching. University of California, Berkeley.

Powell, W.B., 2011. Approximate Dynamic Programming: Solving the curses of dimensionality, 703, 2nd John Wiley & Sons.

Sánchez-Martínez, G., Koutsopoulos, H., Wilson, N., 2016. Real-time holding control for high-frequency transit with dynamics. Transportation Research Part B: Methodological 83, 1–19.

Sun, A., Hickman, M., 2005. The real–time stop–skipping problem. Journal of Intelligent Transportation Systems 9 (2), 91–109.

Wu, W., Liu, R., Jin, W., 2017. Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour. Transportation Research Part B: Methodological 104, 175–197.

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