Transportation Research Part C 105 (2019) 456–467

Contents lists available at ScienceDirect

Transportation Research Part C

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

Optimization-based operations control for public transportation service with transfers

Hend Manasra*, Tomer Toledo

Technion, Israel Institute of Technology, Israel

A B S T R A C T

This research develops a real-time simulation-based control framework that attempts to coordinate the operations of public transportation services to allow smoother transfers and to maintain service regularity. The control actions, which include holding and change speed, are set as the solution of an optimization problem with the objective to minimize total passengers' time in the system within a prediction horizon. The prediction horizon is defined by a number of downstream stops and subsequent buses. The predictions made include the arrival and departure times of vehicles at downstream stops and the passenger demands they are expected to serve. The model is demonstrated with a case study of three lines of the BRT system in Haifa, Israel. The results show that the model outperforms headway control and no-control alternatives in terms of total passengers' time and on-time performance. It is only slightly worst then the headway control in terms of waiting times and denied boarding at stops. The combination of holding and change speed actions provides substantially better results than each of them individually.

1. Introduction

Efficient public transportation systems (PTS) are essential for the provision of sustainable and high-quality mobility in urban areas. Planning agencies worldwide are investing in the design and operation of these systems. However, even an optimally designed PTS may not perform as expected due to service disruptions, road incidents, fluctuations in demand and other unexpected events. In order to realize the potential benefits of PTS to travelers and operators, real-time control of PTS is required. Examples of real-time PTS control include holding, speed change, expressing, stop skipping, short turning and deadheading. All are aimed to maintain regular service and recover from disruptions and unexpected events.

Early works on real-time PTS control treated a single line (e.g., Fu and Yang, 2002; Xuan et al., 2011; Cats et al., 2010, 2011; Eberlein et al., 1999, 2001; Zolfaghari et al., 2004; Cortés et al., 2011; Delgado et al., 2012; Muñoz et al., 2013; Berrebi et al., 2015; Berrebi et al., 2018; Dai et al., 2019). The literature review shows that holding is generally the most effective strategy in term of saving times (Ibarra-Rojas et al., 2015). But, using multiple control strategies jointly may lead greater benefits.

The shift toward design and operation of integrated public transportation services, which require transfers between lines, challenged researchers to develop real time control strategies that deal with multiple lines. Guevara and Donoso (2014) used micro-simulation to study the impacts of stop skipping and offline and online holding on the performance at high demand transfer stops. They developed a rule-based method with the purpose of reducing the variance of bus loads and headways. Dessouky et al. (1999) and Dessouky et al. (2003) developed a simulation system for evaluation of holding strategies. They examined seven different holding strategy combinations on networks that allow transfers. They assumed that passenger demand and bus travel times are deterministic and implemented holding only at transfer stops. Hadas and Ceder (2008, 2010) optimized the total travel time and thereby increased the encounter probability. The strategies they used were holding, stop skipping and slowing down. However, they did not account for the limited capacity of buses. In addition, their optimization function included only passengers' time on the bus and in transfer stops.

* Corresponding author.
E-mail addresses: hend.manasra@gmail.com (H. Manasra), toledo@technion.ac.il (T. Toledo).

https://doi.org/10.1016/j.trc.2019.06.011
Received 24 June 2018; Received in revised form 15 June 2019; Accepted 18 June 2019

Hadas et al. (2013) developed an integer linear programming optimization model that uses holding and stop skipping in order to minimize the total passengers' travel time and maximize the number of direct transfers. However, they did not consider the delays to passengers that are waiting to board a bus at a skipped stop. In addition, the model is myopic in that it assumes that the next bus would be on time and that passenger demands would not change as a result of the control. Liu et al. (2014) optimized the synchronization of planned transfers using holding and change speed. They suggested two different objective functions: minimizing the distance gap between vehicles from the two lines at the transfer stop or the total passengers' travel time in the system. They also defined two types of transfers: at a single transfer point or along a shared corridor with multiple common stops. However, their formulation did not account for delays due to limited vehicle capacities or passengers' waiting time until the arrival of the first bus.

Hernández et al. (2015) evaluated the model of Delgado et al. (2012) to optimize holding for a corridor with multiple bus lines, without transfer stops. They analyzed the level of service of the PT system assuming a central operator that attempts to maximize the level of service for the all system. They developed an optimization model which included in the objective function the waiting time of passengers, the extra waiting time because of limited capacity and due to limited boarding. They examined three different control settings: a central operator that ensures level of service for all passengers in the system, separate operators for each line that are not informed about the position of the competitors' buses, and separate operators for each line that are informed about the position of the competitors buses. They found that a central operator could deliver the best level of service. Nesheli and Ceder (2014) proposed a model that deals with transfer synchronization. A mathematical programming model was formulated to minimize the total passenger travel time and maximize the number of simultaneous transfers with strategies holding and skip certain stops/segments. They showed that considering the skip segment by formulating penalty functions for disadvantaged passengers yielded better results than skipping individual stops. The limitations of their study were mainly related to the assumptions: the passenger demand number and transferring passengers are known and independent of vehicle arrival time, and that the average travel times and dwell times are known.

Later, Nesheli and Ceder (2015) proposed a mathematical programming model to minimize total passenger travel time and maximize direct transfers using mixed integer programming and rolling horizon, with strategies holding, skip stop and short-turn as a real-time control actions. They showed that the combination of the three strategies yield the best benefits in terms of total passengers' time and direct transfers. They also found that the best results are obtained with shorter headways. The limitation of their study is the assumption that the route information, including travel times between stops, estimation of passenger arrival rates at each stop, and average number of transferring passengers are known and fixed over the period concerned.

In summary, few studies focused on real-time control of multiple lines. Some of these studies (e.g. Dessouky et al., 1999, 2003; Hadas and Ceder, 2008, 2010; Guevara and Donoso, 2014; Liu et al. 2014) assumed a single transfer stop and executed control only at that stop. Vehicle capacity constraints were largely ignored (Yap et al., 2019).

This research presents an optimization-based system to control multiple lines. Its objective is to minimize total passenger time including all its components: waiting at stops, travel between stops, dwell times, waiting at transfer stops and additional delays to passengers that are denied boarding. Two control strategies are considered as decision variables in the optimization: holding, which mandates vehicles to dwell at stops longer then they need to board and alight passengers and changing their travel speed in the sections between stops.

The contribution of this research is in the presentation of a comprehensive optimization framework for the control. This framework generalizes previous studies in several ways: First, it considers the entire length of multiple interacting lines, and not only a single transfer stops. Second, it considers all the components of the passengers' trip including riding, waiting, dwelling and transferring times. Finally, it explicitly accounts for vehicle capacities and passengers that are unable to board as a result of these constraints. The optimization model is set in a rolling horizon framework, which may be set to balance between computational complexity of the data acquisition, predictions and optimization and the considering the full effects of control actions.

The remainder of this paper is organized as follows: The next sections present the overall framework for the control system and the details of the optimization problem. This is followed by a case study to demonstrate the system. Finally, discussion and conclusions are presented.

2. Overall system framework

The overall framework is shown in Fig. 1. The optimization process is run whenever a bus enters a stop along its route. The system searches for a set of optimal control actions: holding times at the current and next stops and the travel speed between them. These decisions are estimated for the current bus and for other buses on the same and intersecting lines. In making these decisions it is assumed that the system has information about the current location of all the relevant vehicles. The system then determines the boundaries of the optimization problem to be solved in terms of the buses and stops that would be accounted for in the objective function. The travel times between the stops within this horizon are predicted and used also to predict the passenger demands and dwell times at these stops. These predictions are based on historic and real-time information that may be available. These are used to define an optimization objective to minimize the total passengers' time. Once the optimal solution is found, only the immediate control actions for the current bus at the current stop are implemented: holding and speed change to the next stop. Control actions for subsequent stops and other buses will be re-evaluated when the optimization is run again the next time that a bus enters a stop.

The prediction horizon used in the optimization is defined by a number of stops downstream of the location of the current bus (including the stop that the bus is currently entering), and by a number of buses behind the current bus on the same line (including the current bus), as shown in Fig. 2. Other buses on the same line that are ahead of the current bus, but are still within the stop horizon are also included, and so are buses on other lines that would serve stops within the horizon within the same period of time. The horizon length in the example shown in Fig. 2 is three stops (k, $k+1$, $k+2$) and three buses (i, $i+1$, $i+2$). Bus i on line 1

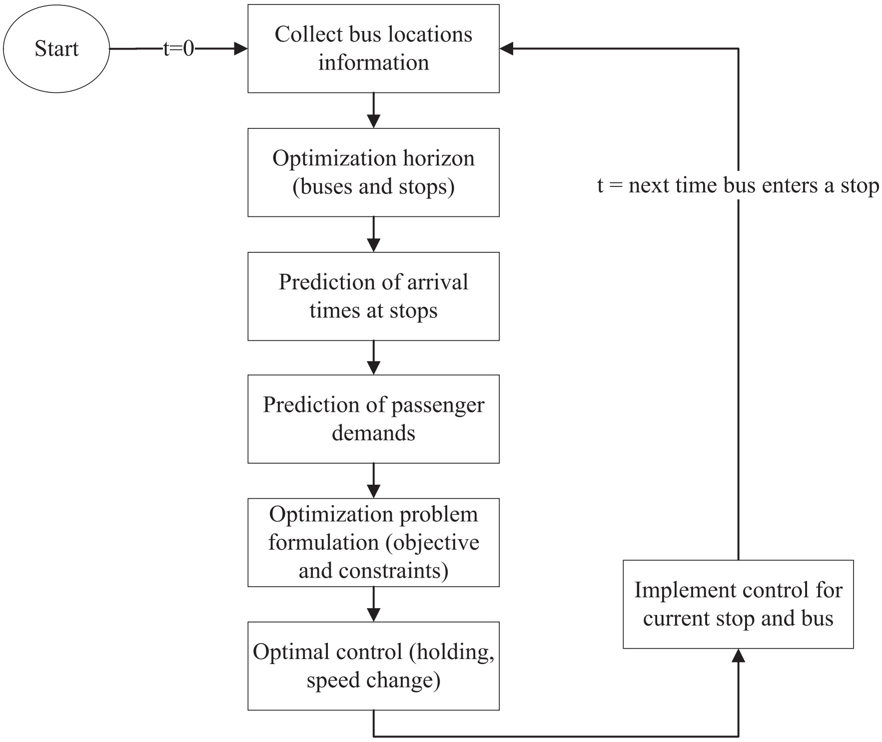

Fig. 1. Overall framework of the optimization system.

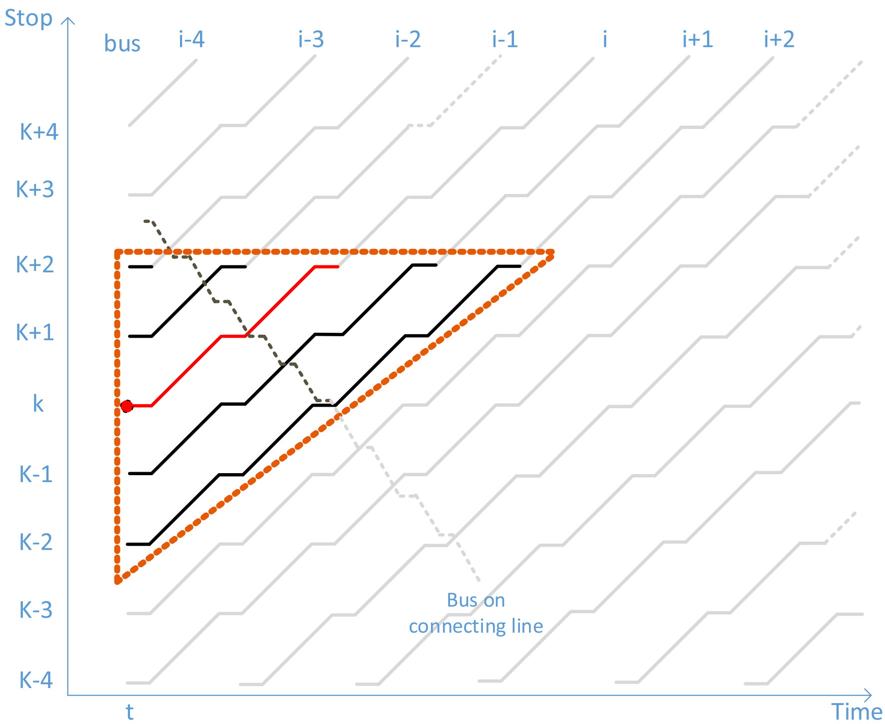

Fig. 2. Rolling horizon of the optimization.

enters stop k and triggers the optimization. The system predicts the arrival time of other buses to that stop. The buses on the same line with the nearest future arrival time to the stop are included in the horizon. Buses that are ahead but still within the prediction horizon (i.e., buses *i-1* and *i-2*) are also included. The optimization objective, and decision variables include these buses at all the stops until the last stop within the horizon. Moreover, if the horizon includes a transfer stop (i.e., stop $k+1$), the system predicts the arrival time of buses on other lines (i.e., line 2) to the transfer stop in order to calculate the waiting time for the transfer passengers, these buses will also be considered in the optimization. Mathematically, the buses and stops in the same line that are included in the horizon are defined by:

$$I_{b,s} = \begin{cases} 1 & \{t(b,s) \geqslant t_0 \cap s < s' + n \cap b < i + m\} \\ 0 & otherwise \end{cases} \tag{1}$$

The buses and stops on the connecting line that are included in the horizon are defined by:

$$I_{b',s'} = \begin{cases} 1 & \{t(b',s') \geqslant t_0 \cap s' \leqslant s^* + n' \cap t(b',s^*) \leqslant t(i+m,s^*)\} \\ 0 & otherwise \end{cases} \tag{2}$$

where $I_{b',s'}$ is an index that equals to 1 if stop s' of bus b' in the connecting line is included in the horizon and 0 otherwise. s^* is the transfer stop, n' is the length of the stops horizon on the connecting line, which may differ from that of the handled line. $t(i+m, s^*)$ is the time that the last bus within the horizon on the handled line arrives at the transfer stop. Thus, the buses on the connecting line that are included in the horizon are those that arrive to the transfer stop before the last bus in the handled line arrive there.

After finding the optimal variable values, the system implements the decisions (holding time and travel speed to the next stop) only for the current bus (i) at the current stop (k). This process is repeated each time a bus enters a stop.

3. Optimization problem

The objective function for the optimization is to minimize the total travel time of passengers within the bus and stop horizon. This includes dwell time at stops (DT), in-vehicle travel time between stops (IVT), waiting time at the origin (WT), waiting time at transfer stops (TWT) and the additional waiting times for passengers that were denied boarding (DBT):

$$Z = \min_{H,TT} \sum_l \sum_b \sum_s \delta_s^{b,l} (\theta_1 DT_s^{b,l} + \theta_2 IVT_s^{b,l} + \theta_3 WT_s^{b,l} + \theta_4 TWT_s^{b,l} + \theta_5 DBT_s^{b,l})$$

where the indices s, b and l signify stops, buses and lines, respectively. $\delta_s^{b,l}$ is an indicator variable that equals to 1 if the specific stop of a specific bus is within the optimization horizon, and 0 otherwise. H and TT are the vectors of decision variables of holding and travel time between stops within the optimization horizon, respectively. θ are parameters that allow to assign different weights to different components of the travel time, which has been shown to exist in passengers' perceptions (Wardman, 2004; Currie, 2005; Garcia-Martinez et al., 2018).

The objective function defined above only includes time components: waiting, travel, dwell and transfer times. However, it can be extended to other level of service measures, such as schedule adherence and crowding levels. For example, in addition to delays, Xiong et al. (2015) include a penalty for crowding conditions on the bus in their optimization objective for timetable and fleet size determination.

The various travel times components within this objective function are calculated using estimates of the numbers of passengers boarding and alighting the buses at the various stops:

Dwell times (DT) are incurred by passengers that are on the bus when it is at a stop boarding and alighting passengers or being held:

$$DT_s^{b,l} = [np_s^{b,l} - (na_s^{b,l} + ntd_s^{b,l} + nta_s^{b,l})](st_s^{b,l} + H_s^{b,l}) \tag{4}$$

where $np_s^{b,l}$ is the number of passengers on the bus when it enters the stop. $na_s^{b,l}$, $ntd_s^{b,l}$ and $nta_s^{b,l}$ are the numbers of alighting passengers: at the destination without and with having made a transfer, and those alighting at the transfer stop, respectively. $st_s^{b,l}$ is the boarding and alighting service time. $H_s^{b,l}$ is the holding time at the stop.

In-vehicle travel times (IVT) are between two consecutive stops. They are incurred by all the passengers on the bus:

$$IVT_s^{b,l} = TT_s^{b,l} np_{s+1}^{b,l} \tag{5}$$

where $TT_s^{b,l}$ is the travel time in the section between stops s and $s+1$.

Waiting times (WT) are incurred by passengers at the origin stop from their arrival at the stop to the departure of the first bus on the line they are served by (regardless of whether or not they are able to board this bus). Assuming that passengers' arrival to the stop is random, the waiting time is given by:

$$WT_s^{b,l} = (nb_s^{b,l} + nto_s^{b,l}) \cdot \left(\frac{dt_s^{b,l} - dt_s^{b-1,l}}{2} \right) \tag{6}$$

where $nb_s^{b,l}$ and $nto_s^{b,l}$ are the number of passengers that want to board the bus at the origin. These are, respectively, those that will only use this line and those that will make a transfer later. $dt_s^{b,l}$ and $dt_s^{b-1,l}$ are the departure times from the stop of the current and previous bus on the same line, respectively.

Transfer waiting time (TWT) is the time that passengers wait at the transfer stop from when they alight the bus they transfer from to the arrival of the bus they need to transfer to:

$$TWT_s^{b,l} = ntb_s^{b,l}(dt_s^{b,l} - at_s^{m,n}) \tag{7}$$

where $ntb_s^{b,l}$ are the number of passengers that want to board the bus at the transfer stop after getting off the first bus on their trip. $at_s^{m,n}$ is the arrival time of the bus the passenger is transferring from to the same stop.

Denied boarding time (DBT) is the additional delay incurred by passengers that cannot board a bus because of crowding.

$$DBT_s^{b,l} = ndb_s^{b-1,l}(dt_s^{b,l} - dt_s^{b-1,l}) \tag{8}$$

where $ndb_s^{b-1,l}$ is the number of passengers boarding the bus that were unable to board the previous bus on the same line due to limited capacity.

Within the calculation of the delays, the various numbers of passengers need to be estimated. The number of passengers on the bus when it enters a stop is calculated based on the load on the bus when it entered the previous stop and the numbers of boarding and alighting passengers in that stop. It is also constrained by the bus capacity:

$$np_s^{b,l} = \min(np_s^{*b,l}, C_{\max}^{b,l}) \tag{9}$$

$$np_s^{*b,l} = np_{s-1}^{b,l} + [nb_{s-1}^{b,l} - na_{s-1}^{b,l} + nto_{s-1}^{b,l} - nta_{s-1}^{b,l} + ntb_{s-1}^{b,l} - ntd_{s-1}^{b,l} + ndb_{s-1}^{b-1,l}] \tag{10}$$

where $np_k^{*s,l}$ is the number of passengers that would like to be on the bus without considering its capacity, $C_{\max}^{s,l}$ is the capacity of the bus.

The numbers of boarding and alighting passengers are estimated using the headways between buses and origin-destination matrices of the direct and transfer demands that are available from historic information. The numbers of passengers waiting to board are given by:

$$n_{sd}^{b,l} = \mu_{sd}^{b,l}(dt_s^{b,l} - dt_s^{b-1,l}) \tag{11}$$

$$nt_{std}^{b,l,n} = \nu_{std}^{b,l,n}(dt_s^{b,l} - dt_s^{b-1,l}) \tag{12}$$

where $n_{sd}^{b,l}$ is the demand of passengers that wish to board bus b at stop s to destination d without making transfers. $nt_{std}^{b,l,n}$ is the demand of passengers that wish to travel on bus b from stop s to destination d making a transfer at stop t to line n. $\mu_{sd}^{b,l}$ and $\nu_{std}^{b,l,n}$ are the corresponding demand rates.

Based on these, the various numbers of passengers are calculated as summation of the relevant passengers' demands (boarding passengers, alighting passengers, transfer boardings at the origin, transfer alighting at the transfer stop, transfer boarding at the transfer stop and transfer alighting at the destination):

$$nb_s^{b,l} = \sum_{d=s+1}^{S} n_{sd}^{b,l} \tag{13}$$

$$na_s^{b,l} = \sum_{j=1}^{s-1} n_{js}^{b,l} \tag{14}$$

$$nto_s^{b,l} = \sum_n \sum_d \sum_{t=s+1}^{S} nt_{std}^{b,l,n} \tag{15}$$

$$nta_s^{b,l} = \sum_n \sum_d \sum_{j=1}^{s-1} nt_{jsd}^{b,l,n} \tag{16}$$

$$ntb_s^{b,l} = \sum_n \sum_m \sum_j \sum_{d=s+1}^{S} nt_{jsd}^{m,n,l}\delta_s^{m,n,b,l} \tag{17}$$

$$ntd_s^{b,l} = \sum_n \sum_m \sum_j \sum_{t=1}^{s-1} nt_{jts}^{m,n,l}\delta_s^{m,n,b,l} \tag{18}$$

where $\delta_s^{m,n,b,l}$ is an indicator variable, which takes the value 1 if bus b is the first bus on line l to serve stop s after bus m on line n serves that stop (i.e. passengers transferring from bus m would use bus b), and 0 otherwise.

The total number of passengers that are unable to board a bus is the excess demand at the stop considering its capacity and current load:

$$ndb_s^{b,l} = np_s^{b,l} + [nb_s^{b,l} - na_s^{b,l} + nto_s^{b,l} - nta_s^{b,l} + ntb_s^{b,l} - ntd_s^{b,l} + ndb_s^{b-1,l}] - C_{\max}^{s,l} \tag{19}$$

If there are denied boarding (the value above is positive), an assumption is needed about which passengers are the ones unable to board. The authors are not aware of literature to guide making an appropriate assumption, which also depends on social norms. On one extreme, a first-in-first-out (FIFO) boarding queue may be assumed. In this case, the last passengers to arrive will not be able to board. At the other extreme, which is the assumption used here, the boarding order is random. Under this assumption, the expected numbers of each type of passengers in the current stop that would be denied boarding are calculated by the proportion factor:

$$\alpha = \frac{ndb_s^{b,l}}{nb_s^{b,l} + nto_s^{b,l} + ntb_s^{b,l} + ndb_s^{b-1,l}} \tag{20}$$

The numbers of passengers that board the bus ($nb_s^{b,l}$, $nto_s^{b,l}$, $ntb_s^{b,l}$, $ndb_s^{b-1,l}$) are adjust down. As a result, the numbers of alighting and transferring passengers in subsequent stops are also adjusted to reflect the actual number of boarding passengers.

The passengers that were denied boarding would need to be served by the next bus. Their total number is given by:

$$ndb_s^{b,l} = ndbb_s^{b,l} + ndbto_s^{b,l} + +ndbtb_s^{b,l} + ndbdb_s^{b-1,l} \tag{21}$$

where $ndbb_s^{b,l}$, $ndbto_s^{b,l}$, $ndbtb_s^{b,l}$ $ndbdb_s^{b-1,l}$ are the denied passengers boarding at the origin, transfer boardings at the origin, transfer boarding at the transfer stop and remaining denied boardings from the previous bus, respectively.

Finally, the various travel times need to be estimated. The departure time from a stop is calculated by the arrival time to that stop and the time that the bus is delayed there:

$$dt_s^{b,l} = at_s^{b,l} + (st_s^{b,l} + H_s^{b,l}) \tag{22}$$

where $at_s^{b,l}$ is the arrival time of the bus to the stop, which in turn is calculated by the departure time from the previous stop and the travel time between the two stops:

$$at_{s+1}^{b,l} = dt_s^{b,l} + TT_s^{b,l} \tag{23}$$

Note that the holding times and travel time between stops affect arrival and departure times and through that the waiting times of passengers at downstream stops.

The service time at a stop is estimated assuming a constant time for each boarding or alighting passenger:

$$st_s^{b,l} = \vartheta_s^{b,l}(na_s^{b,l} + nta_s^{b,l} + ntd_s^{b,l} + nb_s^{b,l} + ntb_s^{b,l} + nto_s^{b,l} + ndb_s^{b-1,l}) \tag{24}$$

where $\vartheta_s^{b,l}$ is the service time for each passenger.

The optimization problem is constrained by bounds on the control variables. The speed change that may be applied is limited (e.g., due to speed limits, traffic conditions). Similarly, the allowed holding time may be bounded. These are captured by the constraints:

$$TT_{s,min}^{b,l} \leqslant TT_s^{b,l} \leqslant TT_{s,max}^{b,l} \tag{25}$$

$$0 \leqslant H_s^{b,l} \leqslant H_{s,max}^{b,l} \tag{26}$$

where $TT_{s,min}^{b,l}$ and $TT_{s,max}^{b,l}$ are the minimum and maximum travel times between the two stops and $H_{s,max}^{b,l}$ is the maximum allowable holding time.

The control system was implemented in C++. The optimization was solved using the L-BGFS-B, a limited memory, box constrained variant of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton algorithm for nonlinear optimization. In every iteration, the algorithm first uses a gradient projection method to solve a quadratic approximation of the objective function to find a set of active bound constraints. The associated variables are fixed to these bounds. The limited quadratic problem with the remaining variables is then solved. The difference between this solution and the solution of the current iteration defines a search direction. A line search is performed along this direction to find the next iteration solution. Finally, the BFGS Hessian approximation is updated. For the full details of the algorithm and its implementation, see Byrd et al. (1995) and Zhu et al. (1997).

4. Case study

4.1. Network

The Metronit BRT network in Haifa, Israel was used for the case study. The network, shown in Fig. 3, is 60 km long, of which 40 km are on dedicated lanes and roadways. It includes three lines with a total of 192 stops, and 19 shared stops. Three main terminals (Hutsot Hamifratz, Lev Hamifratz and Merkazit Hamifratz) are used as transfer stops. They are marked with circles in the figure. In May 2015, the daily ridership on the Metronit was 92,000. 12% of all trips involve a transfer between BRT vehicles.

As shown in Table 1, Line 1 is the busiest in the network. This line is connected to a train station at stop #17 (Merkazit Hamifratz). During the AM peak hour (7:00–8:00) the average headway on this line is 4 min. Fig. 4 shows the profile of line 1 in the case study.

For the case study, travel times and their distributions between stops and dwell times were estimated using automatic vehicle location (AVL) data. Lognormal distributions were fitted for each section separately. Their average coefficient of variation is 0.23. The passengers' and transfer demands were estimated as Poisson distributed with automatic passengers' counter (APC) data and an on-board travel survey conducted by the BRT operator.

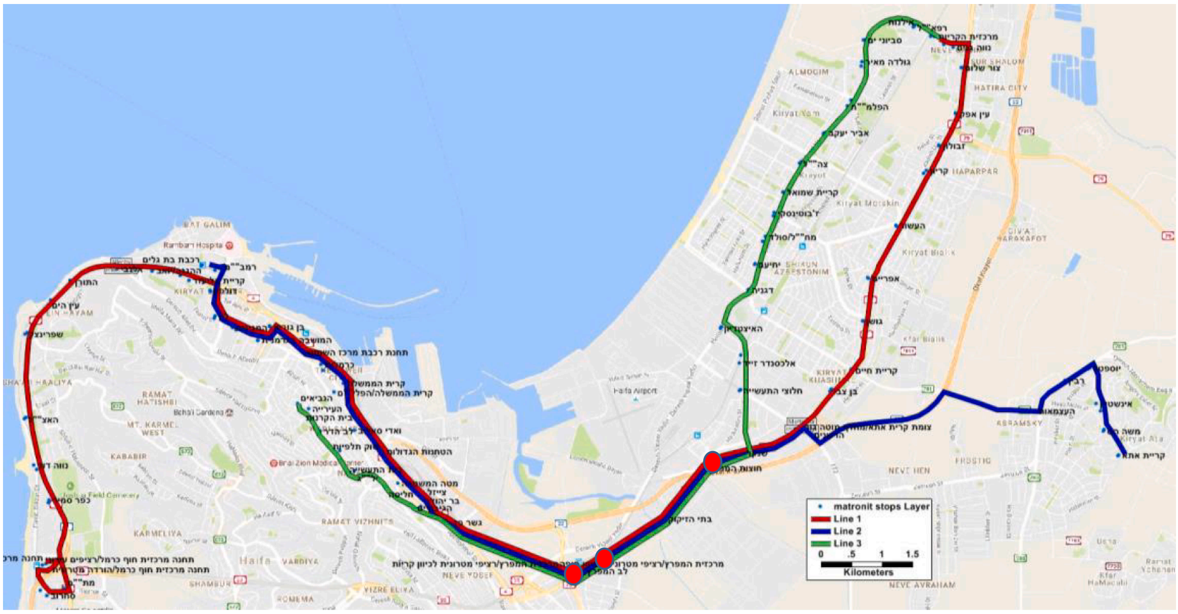

Fig. 3. Metronit network with the main transfer stops.

Table 1
Headways and numbers of passengers in the case study network.

Line	Headway in peak hour (minutes)	Num. passengers in peak hour	Number of stops
1-Northbound	4	2230	41
1-Southbound	4	2600	42
2-Northbound	6	1000	26
2-Southbound	7	420	27
3-Northbound	5	1120	28
3-Southbound	5	650	28

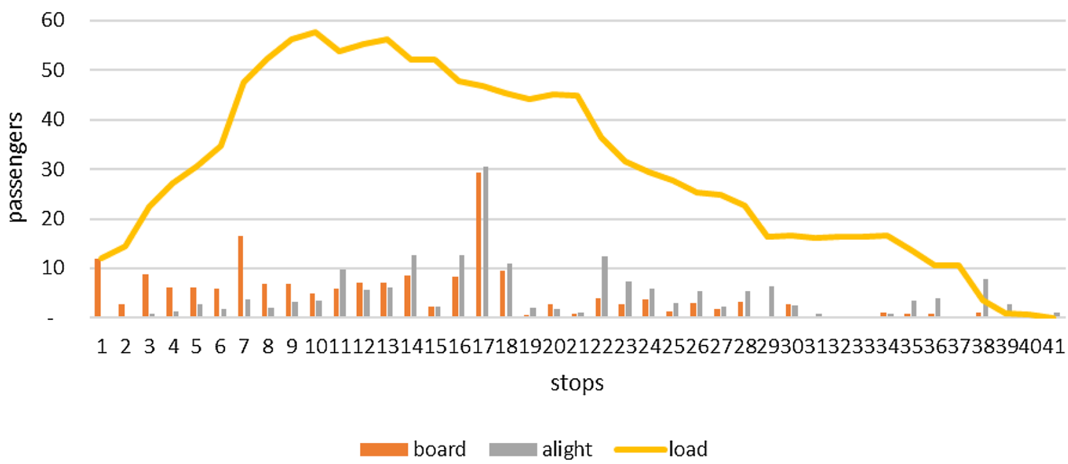

Fig. 4. Demand profile of Line 1-Southbound.

4.2. Scenarios

The performance of the proposed control system is evaluated using BusMezzo (Toledo et al., 2010), a detailed traffic and transit simulation model that generates real-time information that is used within the control systems as well as measures of performance for it. In the case study it was assumed that the information available to the system in real time is only about the current locations of the buses in the BRT network. Predictions of travel times between stops and passengers' demands at stops are based solely on the averages of historic data that were estimated offline. The control system predicts the buses arrival and departure times from stops and the numbers of waiting passengers at these stops. The predicted values used are the historic averages in the AVL and APC data. These are used to define the optimization problem to be solved in order to determine holding times and speed changes. The realized travel times and passengers' demands in the simulation were random draws from their estimated distributions. Therefore, their values are not the same as those used in the optimization.

Holding at a stop was constrained to up to 0.1 of the scheduled headway. The speed change was constrained to be between −5 km/h (slower) and +2 km/h (faster) compared to the mean running speed in the section between two consecutive stops. The implementation of the speed change by drivers depends on traffic conditions, and so drivers may not be able to apply it perfectly. In the simulation, the realized travel times were drawn from a lognormal distribution that was centered on the controlled speed (rather than the historic value) but maintains the historic value of its variance.

For the optimization objective, a weight of two was used for the waiting times (waiting at the origin stop, waiting for transfer and waiting for denied boarding), and 1 for riding times and dwell times. These values are based on estimated perceptions of travel times reported in the literature (Wardman, 2004; Currie, 2005). After some experimentation, as shown in the sensitivity analysis that will be presented later, a horizon of three buses and three stops were used in the optimization.

The proposed optimization control (OC) was compared with a base no control strategy (NC) and a headway-based control (HC) for a single line (implemented on each of the lines, separately). The HC uses holding. It takes into account the headway from both the preceding and the following buses when making holding decisions to a pre-specified minimum headway (Cats et al., 2010):

$$ET_{ijk} = \max\left(\min\left(\begin{array}{c} AT_{ij,k-1} + \frac{(AT_{ijk} - AT_{ij,k-1}) + (AT_{im,k-1} + SRT_{mj} - AT_{ijk})}{2}, \\ AT_{ij,k-1} + \alpha H_i^{k-1,k} \end{array}\right), \begin{array}{c} \\ AT_{ijk} + DT_{ijk} \end{array}\right) \tag{27}$$

where ET_{ijk} is the departure time for the bus on trip k of line i from stop j. AT_{ijk} is the actual arrival time and DT_{ijk} is the dwell time. m is the index of the previous stop that was visited by the bus. SRT_{mj} is the scheduled travel time between stops m and j. $H_i^{k-1,k}$ is the scheduled headway between the two buses. α is a parameter.

The HC was implemented with all stops as time-points. The target minimum headway was set to 80% of the planned headway ($\alpha = 0.8$).

4.3. Results

Table 2 shows the total weighted travel times and its components for the three different control cases (none, headway-based, optimization-based). The total headway-based control has the lowest waiting times. This is not surprising since it is designed to reduce passengers' waiting time. However, this strategy results in much more holding than the other methods. As a result, the total weighted travel time is increased, even compared to the no control case (by 2.1%). The optimization-based control is designed to minimize travel times in the system taking into account all components of the travel time. It is able to reduce the total weighted travel time by 4.0% compared to the no control case. This reduction stems from reductions in all of riding times, dwell times and waiting times, but with much lower holding times compared to the case of headway-based control.

The variability of headways along the southbound of Line 1 are shown in Fig. 5. The coefficients of variation demonstrate that both optimization and headway-based strategies are effective in improving regularity but the effect of the proposed strategy is more pronounced in the second half of the route because it explicitly takes the demand on the line into account.

Table 3 shows measures of performance at the system level. The headway-based control performs best in terms of the number of passengers that are unable to board buses that are full to capacity and waiting times at stops. These results represent decreases of 81% and 17% compared to the no control case on these two performance measures, respectively. The comparable reductions for the optimization-based control are 65% and 13%, respectively. There are no differences in the average waiting time of a passenger that is denied boarding. Waiting times for transferring passengers are reduced by 34% and 11% for the optimization-based and headway-based controls, respectively. The optimization-based control also greatly improves the service reliability. Vehicles are defined as on-time if they arrive between one minute early and three minutes late compared to the scheduled arrival.

The optimization-based control results above used two measures: holding at stops and speed change in sections between stops. The usefulness of each of these measures was evaluated by running optimizations with controls that only implements one of them. Table 4 shows the overall performance of the controls. The two measures used separately yielded low improvements in the total weighted travel times over the no control case: 0.4% and 1.4% for holding and speed change measures, respectively. This is compared to 4.0% for the optimization with both measures. Moreover, the optimization control implementing both measures yielded lower

Table 2
Components of passengers' times (minutes).

Control	Riding time	Dwell time	Total Waiting time	Holding time	Total weighted travel time
None (NC)	212,531	41,667	65,844	0	385,887
Headway (HC)	212,022	39,899	56,798	14,229	393,976
Optimization (OC)	211,018	40,161	58,547	2201	370,475

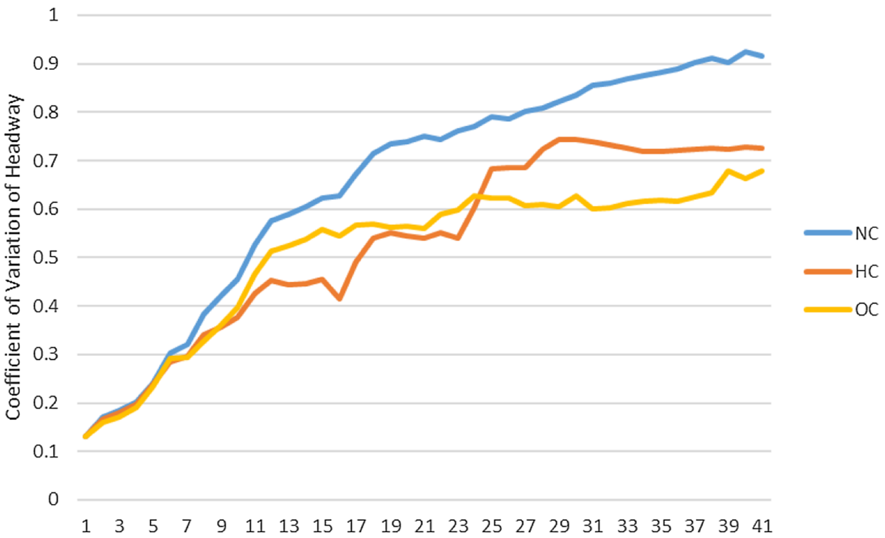

Fig. 5. Coefficient of variation of headways along Line 1.

values in all components of travel time except holding time. It also holds buses less compared to the holding-only strategy. Thus, the results show that the combination of the two measures outperforms using each one of them separately. The speed change measure can compensate for delays caused by holding and so help reduce late arrivals.

Computationally, there were, on average, 40 variables in the optimization in each step. The average running time for the optimization was 1.25 s. These results demonstrate that, at this scale, the L_BGFS-B algorithm performs adequately, and the optimal control system is computationally feasible in real-time.

The choice of the prediction horizons can substantially affect the results. Short bus and stop horizons may not be able to fully capture the negative effects of bunching. They would also lead to later start of the consideration of transfers and their timing within the optimization. Long horizons need to rely on predictions for a longer time period, which are often less accurate. Longer horizons also increase the computational effort associated with the set up and solution of the optimization problems. Results showing the effects of the bus and stop horizons on the total weighted travel times are shown in Fig. 6. The best results are obtained with horizons of three buses and stops. With short horizons, the weighted travel times are longer, even exceeding the no-control case. With longer horizons, travel times increase, but at a lower rate. Computationally, as can be expected, Fig. 7 shows that the optimization time increases with both longer bus and stop horizons. These results suggest that the very long horizons may not be useful, both in terms of the travel time obtained and the computational effort.

5. Conclusion

This study focuses on developing and testing an optimization framework for real-time control of transit systems with transfers. The proposed control method uses two measures: holding and speed change in order to minimize total passengers' travel time that include riding time, dwell time, wait time, transfer time and skip time. The optimization process is implemented in a rolling horizon framework and initiated when a bus enters a stop. The case study implementation demonstrates the computational feasibility of the proposed control system.

The performance of the proposed method is evaluated with a case study of three BRT lines in Haifa, Israel. The simulation-based evaluation uses BusMezzo, a mesoscopic transit simulation model. Two control strategies were used in the case study: holding and speed change. The holding strategy is able promote service regularity and so provides lower waiting and dwell times, but increased riding times and the holding time itself. The speed change strategy reduces riding times but results in increased waiting and dwell times. Their use in combination provided substantially better results compared to each one separately. Applied together, the proposed model reduced the total weighted travel time by 4% compared to the no control case. The average passenger waiting time and the number of passengers that were unable to board the bus decreased by 13% and 65%, respectively. Moreover, the optimization-based control performed better than headway-based control or no control cases, not only in terms of passengers' travel time and its components, but also in terms of service regularity and on-time performance.

The performance of the control depends on the characteristics of the transit service, the passengers demand and control parameters. Relevant service characteristics include the designs of the transit lines, the interactions among them (e.g. transfer points and shared corridors), frequencies and vehicle capacities, use of shared or dedicated lanes and signal priority, variability of travel times and frequency of disruptions in service. Passenger demand factors include the level of demand, its temporal and spatial distribution, share and locations of transfers and the variability of demand. Control parameters include the definition of the optimization horizon, bounds on the allowed control, and weights of the various travel time components. In the case study, sensitivity analysis was conducted for the bus and stop horizon length. In future work, sensitivity analysis for the various factors should be conducted to

Table 3
Service measures of performance.

Control	Passengers unable to board	Average total waiting time (sec)	Average waiting time for denied boarding (sec)	Average waiting time for transfers (sec)	On-time arrivals (%)	Early arrivals (%)	Late arrivals (%)
None (NC)	145	168	47	211	33	56	11
Headway (HC)	28	140	49	190	36	49	14
Optimization (OC)	51	146	49	143	97	2	1

Table 4
Components of passengers' time using a single control strategy (minutes).

Control	Riding time	Dwell time	Waiting time	Holding time	Total weighted travel time
None (NC)	212,531	41,667	65,844	0	385,887
Optimization – Speed change (OSC)	211,287	42,287	65,381	0	384,335
Optimization – Holding (OHC)	213,313	40,786	58,828	4318	380,392
Optimization (OC)	211,018	40,161	58,547	2201	370,475

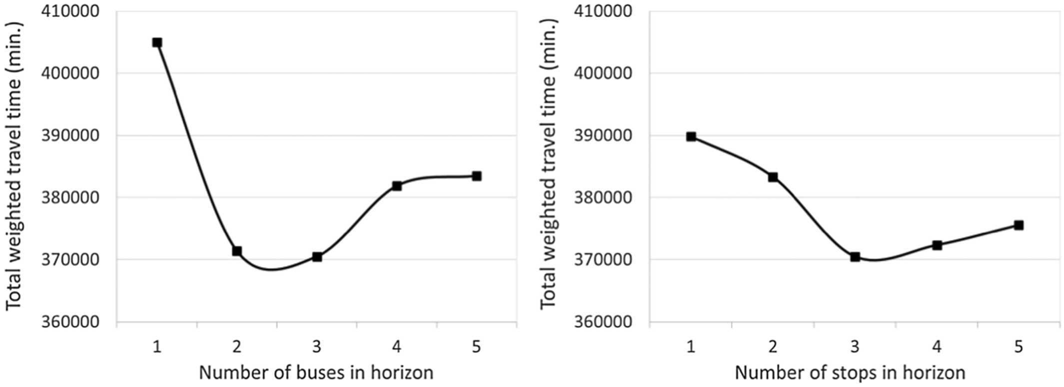

Fig. 6. Total weighted travel times with different bus and stop horizons.

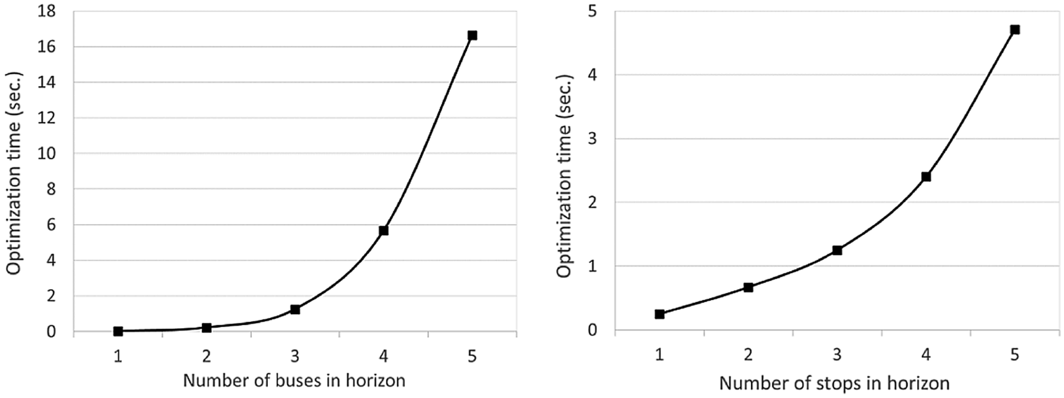

Fig. 7. Optimization computation time with different bus and stop horizons.

identify conditions in which the optimization-based control may be useful, when simpler rule-based control may suffice or when control may not be needed at all.

Implementation of the optimization-based control requires predictions of vehicle travel times between stops and the demand for the transit service in the form of passengers' trip rates between origin to destination stops and their routes within the network that determine transfer stops and rates. In recent years, systems for automatic fare collection (AFC), automatic passenger counters (APC), automatic vehicle location (AVL) systems and smartphone and GPS tracking have emerged that generate large transit data sets. New methods have been proposed that use these data to estimate travel times and the required demand matrices (e.g., Nassir et al., 2011; Gordon et al., 2013; Alsger et al., 2015; Ji et al., 2015). In the case study, these predictions were based only on the means of historic data that was available offline. Recent studies (e.g., Ma et al. 2014; Kieu et al. 2017) show how these estimates may be refined in real-time to reflect actual rather than expected travel times and demand. In future work, these real-time estimates may be incorporated within the control method, which may further improve its performance.

Additional research is also required on other aspects of the data quality, and its availability off-line and in real-time. To be practically useful, the robustness of the method to measurement and prediction errors and its performance in scenarios of service disruptions and surges in demand need to be further studied.

The optimization objective allows to assign different weights for each component in the total passengers' travel time. The choice of weights may affect the control actions. The literature shows that passengers perceive waiting times at the origin and at transfer stops more negatively compared to riding times. Weights that reflect this were used in the case study. Weights can also be used as a policy tool. For example, assigning higher weights to transfer times or denied boarding delays can result in a control that emphasizes transfer coordination or vehicle load balancing and indirectly affect passengers' route choices within the network.

Finally, in the current work, the optimization objective function included only time delay components. Other level of service measures, such as schedule adherence and bus crowding levels may also be used, to reflect the policies and goals of the public transportation agency. Understanding the effect that the choice of objective function has on the service operations and on measures of performance for the system may be helpful to the agencies that make these policy decisions.

Acknowledgement

The authors thank two anonymous referees and the editor for constructive comments and suggestions. Any remaining errors are our own. This research was supported by an ERA-NET grant funded by the Israeli Ministry of Transportation.

References

Alsger, A.A., Mesbah, M., Ferreira, L., Safi, H., 2015. Use of smart card fare data to estimate public transport origin-destination matrix. Transp. Res. Rec. 2535, 88–96.
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. Transp. Res. Part B 81, 377–389.
Berrebi, S.J., Hans, E., Chiabaut, N., Laval, J.A., Leclercq, L., Watkins, K.E., 2018. Comparing bus holding methods with and without real-time predictions. Transp. Res. Part C 87, 197–211.
Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16 (5), 1190–1208.
Cats, O., Burghout, W., Toledo, T., Koutsopoulos, H.N., 2010. Evaluation of real-time holding strategies for improved bus service reliability. In: Proc., 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, Madeira Island, Portugal. IEEE, Piscataway, NJ, pp. 718–723.
Cats, O., Larijani, N., Burghout, W., Koutsopoulos, H.N., 2011. Impacts of holding control strategies on transit performance: a bus simulation model analysis. Transp. Res. Rec. 2584, 51–58.
Cortés, C.E., Jara-Díaz, S., Tirachini, A., 2011. Integrating short turning and deadheading in the optimization of transit services. Transp. Res. Part A: Policy Pract. 45 (5), 419–434.
Currie, G., 2005. The demand performance of bus rapid transit. J. Public Transp. 8 (1), 41–55.
Dai, Z., Liu, X.C., Chen, Z., Guo, R., Ma, X., 2019. A predictive headway-based bus-holding strategy with dynamic control point selection: a cooperative game theory approach. Transp. Res. Part B 125, 29–51.
Delgado, F., Muɒoz, J.C., Giesen, R., 2012. How much can holding and/or limiting boarding improve transit performance? Transp. Res. Part B 46 (9), 1202–1217.
Dessouky, M., Hall, R., Nowroozi, A., Mourikas, K., 1999. Bus dispatching at timed transfer transit stations using bus tracking technology. Transp. Res. Part C 7 (4), 187–208.
Dessouky, M., Hall, R., Zhang, L., Singh, A., 2003. Real-time control of buses for schedule coordination at a terminal. Transp. Res. Part A 37, 145–164.
Eberlein, X.J., Wilson, N.H.M., Bernstein, D., 1999,. Modeling real-time control strategies in public transit operations. In: Wilson, N.H.M. (Ed.), Lecture Note in Economics and Mathematical Systems No. 471: Computer Aided Transit Scheduling. Springer-Verlag, Berlin, Heidelberg, pp. 325–346.
Eberlein, X.J., Wilson, N.H.M., Bernstein, D., 2001. The holding problem with real-time information available. Transp. Sci. 35 (1), 1–18.
Fu, L., Yang, X., 2002. Design and implementation of bus holding control strategies with real time information. Transp. Res. Rec. 1791, 6–12.
Garcia-Martinez, A., Cascajo, R., Diaz, S., Chowdhury, S., Monzon, A., 2018. Transfer penalties in multimodal public transport networks. Transp. Res. Part A 114, 52–66.
Gordon, J., Koutsopoulos, H., Wilson, N., Attanucci, J., 2013. Automated inference of linked transit journeys in London using fare-transaction and vehicle location data. Transp. Res. Rec. 2343, 17–24.
Guevara, C., Donoso, G.A., 2014. Tactical design of high-demand bus transfers. Transp. Policy 32, 16–24.
Hadas, Y., Ceder, A., 2008. Improving bus passenger transfers on road segments through online operational tactics. Transp. Res. Rec. 2072, 101–109.
Hadas, Y., Ceder, A., 2010. Optimal coordination of public-transit vehicles using operational tactics examined by simulation. Transp. Res. Part C 18 (6), 879–895.
Hadas, Y., Ceder, A., McIvorm, M., Ang, A., 2013. Transfer synchronization of public transport networks. Transp. Res. Rec. 2350, 9–16.
Hernández, D., Muñoz, J.C., Giesen, R., Delgado, F., 2015. Analysis of real-time control strategies in a corridor with multiple bus services. Transp. Res. Part B 78, 83–105.
Ibarra-Rojas, O.J., Delgado, F., Giesen, R., Muñoz, J.C., 2015. Planning, operation, and control of bus transport systems: a literature review. Transp. Res. Part B 77, 38–75.
Ji, Y., Mishalani, R.G., McCord, M.R., 2015. Transit passenger origin-destination flow estimation: efficiently combining onboard survey and large automatic passenger count datasets. Transp. Res. Part C 58, 178–192.
Kieu, L.M., Bhaskar, A., Almeida, P.E.M., Sabar, N.R., Chung, E., 2017. Transfer demand prediction for timed transfer coordination in public transport operational control. J. Adv. Transp. 50 (8), 1972–1989.
Liu, T., Ceder, A., Ma, J., Guan, W., 2014. Synchronizing public transport transfers using inter vehicle communication scheme applied to a case study. Transp. Res. Rec. 2417, 78–91.
Ma, Z., Xing, J., Mesbah, M., Ferreira, L., 2014. Predicting short-term bus passenger demand using a pattern hybrid approach. Transp. Res. Part C 39, 148–163.
Muñoz, J.C., Cortés, C., Giesen, R., Sáez, D., Delgado, F., Valencia, F., Cipriano, A., 2013. Comparison of dynamic control strategies for transit operations. Transp. Res. Part C 28, 101–113.
Nassir, N., Khani, A., Lee, S.G., Noh, H., Hickman, M., 2011. Transit stop-level origin-destination estimation through use of transit schedule and automated Data collection system. Transp. Res. Rec. 2263, 140–150.
Nesheli, M.M., Ceder, A., 2014. Optimal combinations of selected tactics for public-transport transfer synchronization. Transp. Res. Part C: Emerg. Technol. 48, 491–504.
Nesheli, M., Ceder, A., 2015. A robust, tactic-based, real-time framework for public-transport transfer synchronization. Transp. Res. Part C: Emerg. Technol. 60, 105–123.
Toledo, T., Cats, O., Burghout, W., Koutsopoulos, H.N., 2010. Mesoscopic simulation for transit operations. Transp. Res. Part C 18 (6), 896–908.
Wardman, M., 2004. Public transport values of time. Transp. Policy 11 (4), 363–377.
Xiong, J., He, J., Guan, W., Ran, B., 2015. Optimal timetable development of community shuttle network with metro stations. Transp. Res. Part C 60, 540–565.
Xuan, Y., Argote, J., Daganzo, C.F., 2011. Dynamic bus holding strategies for schedule reliability: optimal linear control and performance analysis. Transp. Res. Part B 45 (10), 1831–1845.
Yap, M., Luo, D., Cats, O., van Oort, N., Hoogendoorn, S., 2019. Where shall we sync? Clustering passenger flows to identify urban public transport hubs and their key synchronization priorities. Transp. Res. Part C 98, 433–448.
Zhu, C., Byrd, R.H., Lu, P., Nocedal, J., 1997. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization. ACM Trans. Math. Softw. 23 (4), 550–560.
Zolfaghari, S., Azizi, N., Jaber, M., 2004. A model for holding strategy in public transit systems with real time information. Int. J. Transp. Manage. 2, 99–110.