A Novel Model for Transfer Synchronization in Transit Networks and a Lagrangian-based Heuristic Solution Method

Zahra Ansarilari[a], Merve Bodur[b] and Amer Shalaby[a]

[a]Department of Civil and Mineral Engineering, University of Toronto, Canada
[b]Department of Mechanical and Industrial Engineering, University of Toronto, Canada

ARTICLE INFO

Keywords:
Transfer synchronization
Bus dwell time determination
Vehicle capacity
Headway regularity
Mixed integer linear programming
Lagrangian relaxation
Heuristic

Abstract

To realize the benefits of network connectivity in transfer-based transit networks, it is critical to minimize transfer disutility for passengers by synchronizing timetables of intersecting routes. We propose a mixed-integer linear programming timetable synchronization model that incorporates new features, such as dwell time determination and vehicle capacity limit consideration, which have been largely overlooked in the literature on transfer optimization and timetabling problems at the scheduling stage. We introduce a new concept of pre-planned holding time, called transfer buffer time, to reduce the transfer waiting time, particularly for transfers to low-frequency routes, while taking into account the penalty of extra in-vehicle time for onboard passengers and the possible consequences on headway regularity of a route. We develop a Lagrangian relaxation-based heuristic to obtain high-quality solutions efficiently in applications of large instances. Our experiments on instances with up to 12 transfer nodes in the City of Toronto, with a mixture of low- and high-frequency routes, illustrate the potential benefits of the proposed model over a conventional model representing the state of the art found in the literature. The results indicate that incorporating transfer buffer time, dwell time determination, and vehicle capacity limit consideration improves model outcomes considerably. The experiments also demonstrate the computational efficiency of our Lagrangian-based solution method compared to a commercial solver.

1. Introduction

In many transit systems, particularly those with grid designs, a high percentage of passengers have to undertake at least one transfer to complete their transit trips. As strategic elements of transit networks, transfer nodes provide riders with effective connections among pairs of origin-destination locations that are not served by the same line. Arrival and departure times of transit vehicles from different lines meeting at a transfer node should be fully synchronized to enable effective connections. Without well-synchronized route timetables, the inherent disutility of transfers and their role in users' choice of transit (Ceder, 2016) might reduce the perceived attractiveness of transit service and the potential revenues accrued by transit agencies. Therefore, seamless transfers are highly desirable in public transport planning. Transfers occur where two lines are intersecting or terminating at one node. Hence, transfer time is the time interval between alighting from one transit bus (feeder) and boarding another (connecting), which includes walking and waiting times. Although the practical description of a successful transfer is unique, the mathematical formulation of the transfer synchronization problem has varied among previous studies. Several researchers have focused on coordinating the arrival and departure times of feeder and connecting trips through maximizing the number of successful transfers without defining transfer waiting times explicitly (Ceder, Golany and Tal, 2001; Ibarra-Rojas and Rios-Solis, 2012; Ibarra-Rojas, Giesen and Rios-Solis, 2014; Cao, Ceder, Li and Zhang, 2019). On the other hand, some researchers have directly minimized transfer waiting times in the model formulation (Shafahi and Khani, 2010; Abdolmaleki, Masoud and Yin, 2020). Nevertheless, regardless of the type of model formulation, there are some essential modelling considerations concerning the concept of successful transfer, particularly dwell time determination and treatment of bus capacity limit, that previous studies have mostly overlooked while creating timetables.

Most previous studies have considered trip departure times from terminals as the main scheduling decision variables, while timetables at succeeding nodes are determined in a subsequent step using stop-to-stop running times and fixed dwell times. Although using fixed values for various inputs is reasonable and common in deterministic models,

*Corresponding author
zahra.ansarilari@mail.utoronto.ca (Z. Ansarilari)
ORCID(s): 0000-0002-5631-3335 (Z. Ansarilari)

the assumption of a fixed dwell time in timetable design, particularly in transfer synchronization models, may result in unreliable timetables. The components of bus dwell time, namely the service times required by transferring and non-transferring passengers, should be adequately and explicitly considered in transfer synchronization and route timetable models. Otherwise, the generated timetables with fixed dwell times would be prone to fail. For instance, if a group of transferring passengers arrives when a connecting bus is ready to leave based on the schedule with fixed dwell times, these passengers do not have the required boarding time. However, the bus driver in practice would usually wait until all the passengers get on-board. This time interval, which is not considered during timetabling, might be negligible for a bus at one transfer node but cumulatively can lead to long delays in the arrival times of that bus at the subsequent stops. Consequently, designed timetables and planned successful transfers might not occur if a proper bus dwell time calculation is not incorporated at the scheduling stage. The proper treatment of dwell time is also critical when pre-planned holding times are considered in deterministic transfer synchronization models. Similar to the purpose of real-time holding times, some studies have introduced pre-planned holding times incorporated in timetables to increase successful transfer occurrences (Ibarra-Rojas, Muñoz, Giesen and Knapp, 2019) or minimize transfer waiting times (Shafahi and Khani, 2010). However, the omission of dwell time determination from such models and using instead fixed dwell times may negatively impact the final results.

Additionally, in most transfer synchronization models, it is assumed there is sufficient space in connecting vehicles for transferring passengers to board, which is not always true in overcrowded systems. A successful transfer occurs if the connecting bus has enough space for all waiting passengers to board. Therefore, bus capacity consideration in transfer synchronization and timetabling models would provide more realistic results by including the extra waiting time for the next bus experienced by denied passengers, transferring or non-transferring.

This paper aims to fill the aforementioned gaps in the literature by offering a new deterministic formulation for a successful transfer by explicitly considering bus dwell time determination and bus capacity limit. We develop a network-wide transfer synchronization model using a mixed-integer (linear) programming (MIP) approach. We also introduce a new way for incorporating pre-planned holding times in the formulation of the transfer synchronization problem. To illustrate the benefits of our formulation, we benchmark our proposed models, with and without pre-planned holding times, against one conventional transfer time minimization model from the literature and the common practice of route timetabling by agencies in which transfer synchronization is not taken into account. Also, we explore the possible benefits and consequences of including pre-planned holding time in bus dwell time with respect to headway regularity. Moreover, due to the complexity of our transfer synchronization model, we propose an effective solution method. We design a Lagrangian relaxation-based heuristic to obtain high-quality solutions efficiently in applications of large instances. We examine the potential benefits of our solution method through experiments on instances with up to 12 transfer nodes in the City of Toronto, with a mixture of low- and high-frequency routes. Unlike most previous studies, our solution method can provide an optimality gap which provides a quantitative measure of the quality of our solutions.

Note that in this study, the emphasis is placed largely on developing a detailed transfer synchronization formulation under deterministic settings. Ideally, the design of synchronized timetables should take into account the effects of predictable operational variations, such as fluctuations in passenger demand and bus running times. However, it is important to first investigate the deterministic modelling version of the transfer synchronization problem for a number of reasons. First, deterministic models are considerably less computationally demanding than stochastic models, enabling more manageable examination of results. Second, both stochastic and real-time transfer synchronization models are, in fact, extensions of deterministic models since they all have similar objectives and constraints. Third, by developing a detailed deterministic model in which important details of the operation, such as variable dwell times and capacity limit, are considered explicitly, the formulation has the potential to subsequently consider adequately the probabilistic nature of bus dwell times and passenger demand at the stochastic or real-time stages.

The rest of the paper is organized as follows. In Section 2, we review the transfer synchronization literature in detail with respect to the adopted mathematical models and solution methodologies. In Section 3, we provide an overview of our study along with its contributions. In Section 4, we describe the transfer synchronization problem with regards to different types of successful transfers followed by our MIP formulation in Section 5. In Section 6, we present the Lagrangian relaxation-based solution method. In Section 7, we present and discuss the results of numerical examples to highlight different aspects of our mathematical model and its benefits compared to previous approaches. We also elaborate on the practical aspects of the model and the performance of the solution method for a large-scale network in a case study of the City of Toronto, Canada. Finally, in Section 8, we conclude the paper by summarizing our main findings, takeaway messages, and future research recommendations.

2. Literature Review

A large body of literature on the transfer synchronization problem is focused mainly on providing the best dispatching policy for transit lines on fixed routes to generate effective transfer connections. Early researchers have applied different approaches such as interactive graphical optimization (Rapp and Gehner, 1967) and analytical modelling (Wirasinghe, 1980) for the transfer synchronization problem. However, the motivation for implementing mathematical programming approaches that consider different objectives and constraints has increased considerably in recent years. Previous studies either maximize the number of successful transfers or minimize transfer waiting times. Furthermore, several studies have designed various solution methods to efficiently solve transfer synchronization models for real-world large-scale instances. We provide a summary of the related studies in what follows. Note that although there are several approaches for tackling the problem of transfer synchronization, e.g., via stochastic and real-time models, we mostly discuss the research efforts involving deterministic transfer synchronization models since they are closer to the focus of our study.

To maximize the number of successful transfers, Ceder et al. (2001) developed a model to enforce simultaneous arrivals of buses at a transfer node considering each line's minimum and maximum allowable headway while Eranki (2004) defined a secure time window for bus arrivals to improve successful transfers. Later on, researchers have added other features of transit planning in their formulation while maximizing the number of successful transfers. For instance, frequency settings (Zhigang, Jinsheng, Haixing and Wei, 2007; Liu and Ceder, 2018; Silva-Soto and Ibarra-Rojas, 2021), vehicle scheduling (Ibarra-Rojas et al., 2014), as well as bus bunching prevention (Ibarra-Rojas and Rios-Solis, 2012; Ibarra-Rojas, López-Irarragorri and Rios-Solis, 2016) and bus loading consideration (Fouilhoux, Ibarra-Rojas, Kedad-Sidhoum and Rios-Solis, 2016) at transfer nodes have been considered in addition to transfer synchronization.

As mentioned above, another group of transfer synchronization models has focused on minimizing transfer waiting times. Shafahi and Khani (2010) developed a closed-form expression to determine the average transfer waiting times based on the fixed line headways as opposed to the corresponding individual bus transfer waiting time values. The authors used the output of their main transfer waiting time minimization model, i.e., departure times of buses, as input to another MIP model and added pre-planned extra stopping times to fixed dwell times to facilitate better transfers. The explicit consideration of transfer waiting times provides an opportunity to address timetabling problems with more details. For instance, Wong, Yuen, Fung and Leung (2008) minimized transfer waiting times via a MIP model while designing travel times, dwell times, dispatching times, and headways of the transit rail system in Hong Kong. In one study, both transfer waiting times and in-vehicle waiting times have been considered jointly (Tuzun Aksu and Yılmaz, 2014). Another benefit of examining waiting times rather than the number of successful transfer events is to incorporate the passengers' perspective directly into the problem formulation. Schröder and Solchenbach (2006) introduced the concept of "almost transfer", where passengers reach their transfer stop location right after their connecting vehicle leaves. In order to consider a higher penalty of this kind of missed transfer, the researchers categorized transfer events into five types: no transfer, convenience, risky, patience, and almost transfer. The authors stated that although "no transfer" and "almost transfer" are practically the same, i.e., a successful transfer does not occur in both cases, passengers find the latter more bothersome with higher penalty. Therefore, the authors formulated different types of transferring events with different penalties in their quadratic semi-assignment model.

On the other hand, since transfer time is a key influencing factor in passengers' route choice (Ceder, 2016), two studies have developed models to optimize bus timetable synchronization considering passenger route choices, which lead to more realistic demand-based synchronized timetables (Parbo, Nielsen and Prato, 2014; Chu, Korsesthakarn, Hsu and Wu, 2019). In another study, the objective function was to minimize the maximum waiting time while equitably distributing the waiting times over transfer stations in a network (Wu, Liu, Sun, Li, Gao and Wang, 2015).

In the transfer synchronization literature, some studies have also developed multi-objective models. Salzborn (1980) defined the bi-objective problem of minimizing passenger delay and the required number of buses. Ibarra-Rojas et al. (2014) developed a bi-objective model to maximize the number of successfully transferred passengers while minimizing the operational costs related to fleet size. Another study investigated a multi-objective re-synchronization of the bus timetable problem to re-plan the current timetables for better synchronization (Wu, Yang, Tang and Yu, 2016). The objective was to maximize the number of passengers benefiting from transfers and minimizing the maximum deviation from the existing timetables.

Bus capacity limit, particularly for transfer synchronization at the scheduling stage, has been examined only as a decision variable. Liu and Ceder (2018) considered the bus size as a decision variable in a joint model of demand assignment and synchronization of timetables. Also, a decision variable representing the type of a bus was defined in the model aimed at minimizing the observed vehicle load discrepancy as well as the total expected transferring passenger-waiting time (Liu and Ceder, 2016). However, both studies have assumed passengers could board the first connecting bus without trip-by-trip vehicle capacity tracking at the transfer stop. Thus, the models implied no additional delays due to a possible lack of spare capacity, which could undermine model performance.

Maximizing the number of successful transfers and minimizing transfer waiting times have also been applied in a stochastic setting. Gkiotsalitis, Eikenbroek and Cats (2020) developed a robust network-wide transfer synchronization by considering uncertainty in bus travel times and dwell times while maximizing the number of successful transfers for the worst possible scenario. Another study examined the passengers' rerouting due to missed transfers in a stochastic framework in terms of vehicle travel time uncertainty (Wu, Liu, Jin and Ma, 2019). Nonetheless, neither of these studies considered the effect of dwell time distributions and bus capacity limit.

The literature also includes research efforts that aimed at advancing the methodologies of transfer synchronization models. Several researchers have developed procedures and algorithms to reduce their solution search space and efficiently solve their model to optimality. One study defined valid inequalities (Fouilhoux et al., 2016) to strengthen the formulation and reduce the search space. Another study introduced a procedure for reducing the number of decision variables before solving the model. For instance, based on their formulation, they provided the expected range of values for continuous variables and assigned exact values to some binary variables. On the other hand, the literature includes studies that used common solution methods such as local search methods of various types, e.g., hill climbing (Ibarra-Rojas and Rios-Solis, 2012) and genetic algorithm (Shafahi and Khani, 2010; Cao et al., 2019; Ataeian, Solimanpur, Amiripour and Shankar, 2019). Recently, Abdolmaleki et al. (2020) reframed the transfer synchronization as a graph problem on a directed multigraph, and implemented an approximation algorithm for the maximum directed cut problem for its solution. In the studies with multi-objective formulations, successive heuristic algorithms (Liu and Ceder, 2016; Ibarra-Rojas et al., 2019) were designed as the solution approach. For instance, in a study where timetabling was integrated with frequency setting (Ibarra-Rojas et al., 2019), the researchers first determined the frequencies of lines, and subsequently, by fixing passenger-route assignment based on the obtained frequencies, they solved the timetabling problem via a commercial solver.

3. Contributions

Previous studies on transfer optimization have mostly formulated their transfer time minimization models with a focus on improving the synchronization of feeder and connecting buses' arrival/departure times without considering the consequences of using fixed dwell time and the possibility of capacity limitations of the connecting vehicles. To address these shortcomings, our study contributes to this area by proposing a new formulation of the transfer optimization problem with and without the inclusion of pre-planned holding time as determined by an explicit dwell time formulation and considering the spare capacity available for passengers to board. The contributions of this work to the state of the art are:

- new formulation of a successful transfer process with and without pre-planned holding time based on transferring passengers' arrival time to their connecting stop;
- new formulation of bus dwell time model considering the number of successful boarding passengers, transferring and non-transferring, based on vehicles' spare capacity and pre-planned holding time;
- assignment of pre-planned holding time to accommodate transferring passengers while taking into account the cost of extra holding time to passengers already on-board;
- consideration of bus capacity in a transfer synchronization model at the scheduling stage;
- examination of our model performance compared to a common transfer synchronization model in the literature; and
- introduction of a Lagrangian-based heuristic solution method that provides lower and upper bounds on the optimal objective value.

In what follows, we explain each of the contributions in detail. First, we define two new concepts in this study, namely "dwell transfer buffer" and "transfer buffer" times. The dwell transfer buffer time can indicate different possible arrival times of transferring passengers at their connecting bus stop location, i.e., before or after the connecting bus arrival. This differentiation helps us determine passengers' transfer waiting times and their required boarding time. If transferring passengers arrive before the arrival of the connecting bus, they have to wait. However, if they arrive while the connecting bus is already at the stop, they do not need to wait. On the other hand, the transfer buffer time is a pre-planned holding time of a bus that has been implemented in previous models (Shafahi and Khani, 2010; Voorhorst and Gkiotsalitis, 2020) in which critical details have been overlooked. In our model, the transfer buffer time is calculated more accurately by taking into account the dwell time determination of a bus and its capacity limit. In particular, we investigate how the added transfer buffer times affect a transit system from two different perspectives, namely that of passengers and and with respect to bus operation.

From the passengers' perspective, by adding extra time to a bus dwell time, the passengers already on-board experience longer in-vehicle time. To consider this cost, Voorhorst and Gkiotsalitis (2020) have defined a penalty as a coefficient in their objective function. However, the authors have not explicitly considered the number of passengers who are penalized. In our study, we introduce the transfer buffer time to minimize the total transferring passenger waiting times while considering the extra penalty of in-vehicle time based on the number of passengers already on-board. Therefore, the model has the opportunity to quantify the on-board passengers' in-vehicle cost and compare it with the time saving of transferring passengers who can board their connecting bus without incurring extra waiting time.

With respect to the bus operation, it is important to consider these additional holding and boarding times in the bus dwell time calculation. Not only should arrival times of buses at preceding nodes be adjusted correctly, but also the associated bus cycle times should be calculated more accurately. Therefore, we determine the required dwell time of a bus that includes service time for the alighting and boarding of transferring and non-transferring passengers as well as the assigned transfer buffer time. Thus, the explicit estimation of dwell time should improve the effectiveness of timetables in practice.

Moreover, since the dwell time computation depends on the number of boarding and alighting passengers, the bus capacity limit needs to be observed. We use the same formulation of our previous work (Ansarilari, Nesheli, Srikukenthiran, Bodur and Shalaby, 2018) for detecting leftover passengers of buses due to capacity limitations. As mentioned, adding transfer buffer time is reasonable when the bus has enough space for the incoming passengers; otherwise, there is no point in holding the bus more than its dwell time unless it is used for headway regularity purposes. To the best of our knowledge, the formulation of exact dwell time while considering capacity limit has not been developed previously at the scheduling stage.

By incorporating bus dwell time determination and bus capacity constraints, our model becomes computationally difficult to solve by state-of-the-art MIP solvers such as CPLEX and GUROBI. Thus, it is vital to have an efficient method to derive high-quality solutions, if not optimal, to the model within an acceptable duration for a large-scale network. We design a Lagrangian relaxation-based heuristic in which the main network model is broken down into separate transfer node-based optimization models, called subproblems. The Lagrangian relaxation yields a lower bound on the optimal objective value of our proposed MIP model. However, the solution obtained from the relaxation might not be feasible to the original MIP model. Therefore, based on the solution of this lower-bound problem, we apply another heuristic to produce a feasible solution to our problem, as such an upper bound on the optimal value. Generating lower and upper bounds is repeated over a number of iterations, during which the Lagrangian multipliers are updated based on the best achieved bounds. Finally, when an acceptable optimality gap is achieved, or the Lagrangian dual problem is solved to optimality, we can either output the best discovered feasible solution as the final solution, or in order to further improve it we use it as a warm-start solution to our MIP model and solve it with a commercial MIP solver. In this approach, there is a possibility of finding high-quality solutions in a reasonable computational effort, rather than spending a prohibitive time to achieve an optimal one. Another benefit of our approach is that the model complexity is reduced thanks to the decomposition, as such it can scale well with the number of transfer nodes, which represents a significant improvement over the previous studies.

As discussed above, most previous contributions in terms of solution methods have proposed various heuristics and hybrid solution methods to render the developed models efficiently solvable for large transit networks. Nevertheless, in most cases, the offered methods were designed to provide solutions without a clear indication of how far they are from an optimal solution. Such an approach may not be desirable when complex models are applied for large-scale networks. However, in this study, we develop a solution method to provide an optimality gap while solving the model towards optimality. Thus, unlike previous methods, our solution method guarantees the quality of the obtained solutions, i.e., by providing lower and upper bounds throughout. In many cases, transit agencies might prefer good-quality solutions over an optimal one that might require much longer computational time.

4. Problem Definition

Consider a transit network with a pre-determined set of transfer nodes N and set of lines L. The network layout, frequency of each line, and bus size have been determined beforehand and are considered to be fixed during the given planning horizon. In this study, we seek to minimize transfer waiting times between bus lines at the chosen transfer nodes for all possible transfer directions by deciding on the arrival times of buses at the sequence of transfer nodes, their associated dwell times including transfer buffer time, and departure times. In what follows, prior to presenting our MIP model, we discuss the model's assumptions, and we explain in detail three possible situations for a successful transfer.

Note that our formulation for dwell time and capacity limit consideration is based on detailed passenger demand. With the growing availability of detailed sensor data such as AVL (Automatic Vehicle Location), APC (Automatic Passenger Count), AFC (Automatic Fare Collection), and WiFi detailed measurements of passenger demand by location can be obtained. Accordingly, the assumptions made in this study are as follows:

- Line headways are given; however, they are flexible based on a specified given range of minimum and maximum allowable headway (h_l^{min} and h_l^{max} for line l) chosen by the user.
- For each planning horizon (e.g., morning peak period) bus running times between transfer nodes are given.
- The number of local passengers coming from streets, alighting passengers from a bus, and transferring passengers in each feeder bus at transfer nodes as well as the total net number of passengers alighting and boarding at the stops between the selected transfer nodes are given for each bus at all transfer nodes.
- The number of in-vehicle passengers are given only for the buses of each line at their first transfer node. (For the other nodes, these values are calculated in the model based on the previous nodes' alightings and boardings.)
- Passengers' walking time between stops (from a feeder bus to a connecting bus) is given.
- Although passengers' arrival patterns depend on many factors such as a line's headway, local passengers arrive at the stop before the arrival time of their bus.
- The boarding and alighting doors are separate.

As mentioned above, one of our contributions is the new approach of successful transfer formulation using two concepts of dwell transfer buffer (DTB) and transfer buffer (TB) times. We distinguish between three types of possible successful transfers to a connecting bus which are shown in Figure 1 and will be explained in detail below. Note that with the new concepts of DTB and TB, the problem gets more complicated regarding capacity limit consideration. Various situations due to lack of capacity can happen for each type of successful transfer. Thus, we check the capacity limit for all three types of successful transfers separately to determine the number of actual boarding later in Section 5. Therefore, to clearly explain the three types of a successful transfer, in the following description, the assumption is that all passengers can board their bus without any capacity limitation.

Type 1 of successful transfer: transferring passengers arrive before the arrival of a connecting bus

Successful transfer Type 1 occurs if a group of passengers arrives at their connecting bus stop location before the connecting bus arrival. The associated transfer waiting time is the difference between the passengers' arrival time at the stop location and the arrival time of the connecting bus. Accordingly, the first component of the bus dwell time is the service time of Type 1 successfully transferred passengers and local passengers coming from the streets, `serv1`.

Type 2 of successful transfer: transferring passengers arrive during `serv1`

In Type 2 of successful transfers, a group of transferring passengers arrives during `serv1`. In this case, they have to wait until the first group from feeder 1 boards the connecting bus. Note that, feeder 1 and feeder 2 are trips from different lines. To distinguish this type of successful transfer, the `dtb` interval is defined with a value less than the first service time, `serv1`. The `dtb` interval helps us indicate the transferring passengers who arrive at the connecting stop location after the arrival of the connecting bus while getting a successful transfer without any extra holding time of the connecting vehicle. In other words, it shows that this group of transferring passengers arrive when the bus is in service, i.e., no holding time. Thus, there is no penalty for in-vehicle passengers, unlike successful transfer Type 3,

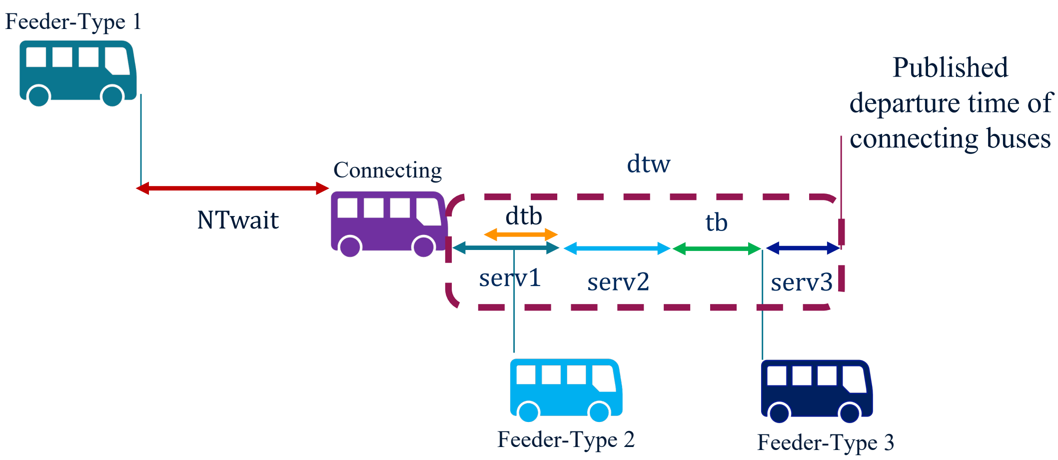

Figure 1: The process of three types of a successful transfer.

which will be explained later. The service time of the second group, denoted by `serv2`, starts right after `serv1`. Its value is determined based on the number of Type 2 successfully transferred passengers who can board, considering the bus capacity. If `serv1` equals zero, `dtb` is also zero, so passengers arrive at the same time as their connecting bus. This condition is similar to what happens in a Transfer Time System (TTS) design when buses' departure or arrival times are set to be simultaneous. This type of successful transfer is preferable for both passengers and agencies since there is no waiting time for passengers and no cost of extra holding time for agencies.

Type 3 of successful transfer: with the help of `tb`

The third type of a successful transfer happens with the help of `tb`. In this type, we add a transfer buffer time to hold the bus after the boarding time of previous passengers, `serv1` and `serv2`. The `tb` time indicates holding a bus without any service, i.e., no boarding, to allow another group of transferring passengers to get to this connecting bus without waiting for almost another headway for the next bus. Nevertheless, this group of passengers also needs their own boarding time, denoted by `serv3`. If this group of passengers arrives during or right after the service time of the second successfully transferred passengers, `serv2`, `tb` equals zero; it means that there is no holding assigned to a bus. Otherwise, the extra time, including `tb` and `serv3`, should be considered as an in-vehicle time penalty for the passengers already on-board. These passengers include the ones who were on the bus from previous stops of the connecting line (i.e., through passengers), the transferring passengers who get on this connecting bus via Type 1 and Type 2, and the passengers who get to this bus from local streets. Note that the value of `tb` can have some restrictions applied by agencies which are discussed later.

As mentioned earlier, although previous studies (Shafahi and Khani, 2010; Voorhorst and Gkiotsalitis, 2020) have examined the TB concept, they did not determine the exact dwell time. Therefore, their formulation could not recognize and take into account the consequences of adding extra time to a bus holding time. In our study, regardless of allowing transfer buffer time, we can determine a bus dwell time while considering the associated costs and benefits of transfer buffer times as well as bus capacity limit. Therefore, our formulation would lead to more reliable arrival and departure times of buses compared to previous models if utilized in practice.

5. Mathematical Formulation: MIP Model

In this section, we first introduce the notation used in our formulation. Then, we explain the details of the objective function and constraints of our proposed MIP model. The sets and parameters are introduced in Table 1. The time-related and demand-related variables are presented in Table 2 and Table 3, respectively. All the variables are non-negative and continuous unless otherwise stated. Note that we use a few notational conventions to improve understanding and legibility. We use upright letters for decision variables while indicating sets, indices, and parameters

Table 1
Notation: Sets and parameters

Sets:	
N	Set of chosen transfer nodes of the whole network; $n \in N$
L	Set of lines passing at least one of the chosen transfer nodes in the network; $l \in L$
TN_l	Set of transfer nodes along line $l \in L$; $n \in TN_l$
TP	Set of transfer pairs of lines at transfer nodes; $(n, l, l') \in TP$, $n \in TN_l \cap TN_{l'}$, $l, l' \in L$, $l \neq l'$
CL_n	Set of connecting lines at transfer node $n \in N$; $l' \in CL_n$
P_l	Index set of feeder buses of line $l \in L$ during the planning horizon; $p \in P_l = \{1, 2, \dots, \lfloor T/h_l \rfloor\}$
$Q_{l'}$	Index set of connecting buses of line $l' \in L$ during the planning horizon; $q \in Q_{l'} = \{1, 2, \dots, \lfloor T/h_{l'} \rfloor\}$
$Q^-_{l'}$	Index set of connecting buses of line l' during the planning horizon without the last bus; $q \in Q^-_{l'} = \{1, 2, \dots, \lfloor T/h_{l'} \rfloor - 1\}$;
$Q^*_{l'}$	Index set of connecting buses of line $l' \in L$ during the planning horizon, $Q_{l'}$, and the first trip of the next planning horizon; $q \in Q^*_{l'} = Q_{l'} \cup \{q^*_{l'}\}$
Parameters:	
T	Length of the planning time horizon
h_l	Headway of line l during the planning horizon; $l \in L$
h_l^{min}	Minimum allowable headway of line l during the planning horizon; $l \in L$
h_l^{max}	Maximum allowable headway of line l during the planning horizon; $l \in L$
$Mtb^n_{l'}$	Maximum allowable transfer buffer time for line l' at each transfer node during the planning horizon; $l' \in L$
$Ttb^{max}_{l'}$	Total maximum allowable transfer buffer time for line l' across all nodes during the planning horizon; $l' \in L$
$n_1(l)$	The first transfer node index of line l; $l \in L$
n_{prev}	The transfer node prior to node n in the selected nodes sequence of a line, or terminal if n is the first node of the line; $l \in L$, $n_{prev} \in TN_l$
$ttd_l^{n_{prev}-n}$	Travel time distance between two consecutive transfer nodes, or terminal to the first transfer node, for line l from the previous node n_{prev} to its consecutive one n in the sequence; $l \in L$, $n \in TN_l$
$awt^n_{ll'}$	Average walking time from stop locations of line l and l' at node n; $(n, l, l') \in TP$
$td^n_{lpl'}$	Number of transferring passengers in vehicle p of line l who want to transfer to line l' at node n; $(n, l, l') \in TP$, $p \in P_l$
$ld^n_{l'q}$	Number of local boarding passengers of bus q of line l' at node n; $l' \in L$, $q \in Q_{l'}$, $n \in TN_{l'}$
$ad^n_{l'q}$	Number of alighting passengers from bus q of line l' at node n; $l' \in L$, $q \in Q_{l'}$, $n \in TN_{l'}$
$sp^{n_{prev}-n}_{l'q}$	Net number of people alighting and boarding at the stops between the two consecutive selected transfer nodes n_{prev} and n for bus q of line l'; $l' \in L$, $q \in Q_{l'}$, $n_{prev}, n \in TN_{l'}$
$vcn_{l'q}$	The nominal bus capacity of bus q of line l'; $l' \in L$, $q \in Q_{l'}$
$q^*_{l'}$	The first trip of line l' in the next planning horizon; $l' \in L$
c^{tw}	Relative penalty cost of transfer waiting time
c^{fl}	Relative penalty cost of passengers who miss their first connecting bus due to lack of capacity
c^{sl}	Relative penalty cost of leftover passengers who miss their second connecting bus due to lack of capacity
c^{vt}	Relative penalty cost of extra in-vehicle time for in-vehicle passengers
c^h	Relative penalty cost of headway irregularity for local passengers
c^b	Relative penalty cost of bunching for local passengers
a^t	Alighting time required per passenger
b^t	Boarding time required per passenger
M_i	Large (big-M) numbers used to formulate logical constraints; $i = 1, 2, \dots$

with the italic font. Also, in transfer-related components l and p represent indices related to a feeder bus, whereas q and l' refer to a connecting bus. Nevertheless, in general cases, such as arrival and departure times of buses, l and p represent indices related to both feeder and connecting buses. On the other hand, if only q and l' indices are used, it refers solely to connecting buses.

The objective function consists of four components. The first component, (1a), refers to transferring passengers

Table 2
Notation: Time-related decision variables

firstarr_l	Arrival time of the first bus of line l at its first transfer node, $l \in L$
For each $n \in N, l' \in CL_n, q \in Q_{l'}$:	
$\text{arr}^n_{l'q}$	Arrival time of bus q of line l' at node n
$\text{dep}^n_{l'q}$	Departure time of bus q of line l' from node n
$\text{dwt}^n_{l'q}$	Total dwell time of bus q of line l' at node n
$\text{serv1}^n_{l'q}$	Service time for local passengers and Type 1 of successfully transferred passenger to bus q of line l' at node n
$\text{dtb}^n_{l'q}$	Dwell transfer buffer time of bus q of line l' at node n
$\text{serv2}^n_{l'q}$	Service time of bus q of line l' at node n added for successfully transferred passenger with the help of dwell transfer buffer time, Type 2
$\text{tb}^n_{l'q}$	Transfer buffer time for bus q of line l' at node n (can be negative)
$\text{ptb}^n_{l'q}$	Non-negative transfer buffer time for bus q of line l' at node n (equal $\text{tb}^n_{l'q}$ if it is positive)
$\text{Ib}^n_{l'q}$	0: if $\text{ptb}^n_{l'q}$ is zero, 1 otherwise; for bus q of line l' at node n
$\text{serv3}^n_{l'q}$	Service time of vehicle q of line l' at node n added for successfully transferred passenger with the help of transfer buffer time, Type 3
$\text{TI}^n_{l'q}$	1: if service time needed for boarding passengers and $\text{ptb}^n_{l'q}$ is more than the time needed for alighting passengers for bus q of line l' at node n, 0: otherwise
$hreg^n_{l'q}$	The quantity of headway regularity violation for bus q of line l' at node n
$bun^n_{l'q}$	The time difference between arrival of a bus at a node and the departure of previous bus in that line for vehicle q of line l' at node n
For each $(n,l,l') \in TP, p \in P_l, q \in Q^*_{l'}$:	
$\text{Y}^n_{lpl'q}$	1: if arrival time of transferring passengers from bus p of line l is less than arrival time of bus q of line l' at node n plus $\text{dtb}^n_{l'q}$, 0: otherwise
$\text{ZY}^n_{lpl'q}$	1: the first eligible connecting bus q of line l' for transferring passengers of feeder vehicle p of line l at node n; 0: either not eligible or not the first eligible connecting bus
$\text{SY}^n_{lpl'}$	1: if there is a successful transfer of Type 1 or 2 from feeder bus p of line l to connecting line l' 0: if there is a successful transfer of Type 3 from feeder bus p of line l to connecting line l'
$\text{TY}^n_{lpl'q}$	1: if there is a successful transfer, Type 1 or 2 from feeder bus p of line l to connecting bus q of line l' at node n, 0: otherwise
$\text{YY}^n_{lpl'q}$	1: if there is a successful transfer of Type 3 from feeder bus p of line l to connecting bus q of line l' at node n, 0: otherwise
$\text{NTwait}^n_{lpl'q}$	Transfer waiting time of passengers from bus p of line l transferring to bus q of line l' at node n , i.e., the exact time people wait for their connecting bus to arrive (if passengers arrive during the service time of bus q this value is zero)
$\text{Twait}^n_{lpl'q}$	Transfer waiting time of passengers from bus p of line l transferring to bus q of line l' at node n with the help of dwell transfer buffer time
$\text{Iwait}^n_{lpl'q}$	1: if the difference of $\text{NTwait}^n_{lpl'q}$ and $\text{Twait}^n_{lpl'q}$ is exactly equal to $\text{dtb}^n_{l'q}$, 0: otherwise

and their transfer waiting time. The second component, (1b), is defined to consider the penalty of extra in-vehicle time for passengers who are from the previous stops (i.e., through passengers), local passengers, and the transferring passengers who can board their connecting bus without the help of `tb`. In order to prevent non-linearity in the model, we do not consider the duration of the transfer buffer time in the objective function; instead, we assume a fixed penalty for passengers who experience extra transfer buffer times. They would experience the maximum allowable transfer buffer at the longest, for instance, 10% of a line's headway. Although the actual extra in-vehicle time includes `serv3` but in most cases, `serv3` is negligible compared to the maximum allowable transfer buffer. The third component, (1c), takes into account the penalty of transferring and/or non-transferring passengers who miss their first assigned bus due to the lack of capacity, for which we consider the connecting line's headway as the extra waiting time for these first leftover passengers. Similarly, (1d) refers to the penalty of passengers who miss their first assigned connecting bus and wait for the next bus but missed that one as well due to the lack of capacity. For those second leftover passengers in

Table 3
Notation: Demand-related decision variables

For each $n \in N, l' \in CL_n$, $q \in Q_{l'}$	
$\text{ivd}^n_{l'q}$	Number of in-vehicle passengers of vehicle q of line l' when it arrives at node n
$\text{EC}^n_{l'q}$	Available capacity of bus q of line l' at node n
$\text{GBD1}^n_{l'q}$	Total number of local and transferring passengers who get successful transfer of Type 1 to bus q of line l' at node n
$\text{Tbd1}^n_{l'q}$	Total number of passengers from $\text{GBD1}^n_{l'q}$, who are able to board on bus q of line l' at node n
$\text{GBD2}^n_{l'q}$	Total number of transferring passengers who get successful transfer of type two to bus q of line l' at node n
$\text{Tbd2}^n_{l'q}$	Total number of passenger from $\text{GBD2}^n_{l'q}$, who are able to board on bus q of line l' at node n
$\text{Ttd}^n_{l'q}$	Total number of successfully transferred passengers, all types, whose first possible connection is bus q of line l' at node n
$\text{Tbd}^n_{l'q}$	Total number of passenger who are able to board on bus q of line l' at node n
$\text{CI}^n_{l'q}$	1: if the sum of the total demand, i.e., $\text{Ttd}^n_{l'q}$, and leftover passengers, for bus q of line l' at node n is less than the existing capacity of the bus, 0: otherwise
$\text{CI1}^n_{l'q}$	1: if the $\text{GBD1}^n_{l'q} + \text{flp}^n_{l'(q-1)}$ demand for vehicle q of line l' at node n is less than the existing capacity of the bus, 0: otherwise
$\text{CI2}^n_{l'q}$	1: if the $\text{GBD2}^n_{l'q}$ demand for bus q of line l' at node n is less than the existing capacity of the bus, 0: otherwise
$\text{pd}^n_{l'q}$	Number of passengers getting their connecting bus without the help of transfer buffer plus local passengers, i.e., penalized passengers if being held due to positive transfer buffer while no alighting service occurs either, for bus q of line l' at node n
$\text{flp}^n_{l'q}$	Total number of local and transferring passengers who miss their first eligible bus which is bus q of line l' at node n due to lack of capacity
$\text{slp}^n_{l'q}$	Total number of local passengers and transferring passengers who have missed their first eligible bus which is bus q of line l' at node n and also missed their second eligible bus due to lack of capacity (thus assumed to leave the system)
$\text{FL}^n_{l'q}$	1: if the existing capacity is less than the generated demand so $\text{flp}^n_{l'q}$ is positive for bus q of line l' at node n due to lack of capacity; 0: otherwise
$\text{SL}^n_{l'q}$	1: if the existing capacity is less than the number of first leftover passenger of previous bus, so $\text{slp}^n_{l'(q-1)}$ is positive, for bus $(q-1)$ of line l' at node n due to lack of capacity; 0: otherwise

(1d), two times of a line's headway is applied. Each component is also weighted with an appropriate cost coefficient, which can be chosen based on the related literature or agencies' recommendations to capture people's perception of different waiting times.

$$\min \sum_{(n,l,l')\in TP} \sum_{p\in P_l} \sum_{q\in Q^*_{l'}} td^n_{lpl'}(\text{NTwait}^n_{lpl'q})c^{tw} + \tag{1a}$$

$$\sum_{n\in N} \sum_{l'\in CL_n} \sum_{q\in Q_{l'}} \text{pd}^n_{l'q}(Mtb^n_{l'})c^{vt} + \tag{1b}$$

$$\sum_{n\in N} \sum_{l'\in CL_n} \sum_{q\in Q_{l'}} \text{flp}^n_{l'q}(h_{l'})c^{fl} + \tag{1c}$$

$$\sum_{n\in N} \sum_{l'\in CL_n} \sum_{q\in Q_{l'}} \text{slp}^n_{l'q}(2h_{l'})c^{sl} \tag{1d}$$

The arrival time of each bus at the first transfer node of their sequence is one of the primary decision variables. Accordingly, a trip's departure time from its starting terminal can be determined based on the running time between the line's first transfer node and its terminal and the recommended arrival time at its first transfer node. Our model focuses directly on optimizing transfer waiting times during a specified planning horizon. Note that although the planning horizon that the model aims to synchronize the trips at selected transfer nodes is the same for all lines, for example, 6-9 am, each line can have a unique operation horizon. In other words, there is no need for all lines to begin their trips from route terminals based on the same starting time, e.g., 6 am. A line may depart from a terminal at 5:30 am and another at 6:10 am.

$$\text{arr}^n_{lp} \leq h^{max}_l \qquad l \in L, n = n_1(l), p = 1 \tag{2}$$

$$\text{arr}^n_{lp} - \text{arr}^n_{l(p-1)} \geq h^{min}_l \qquad l \in L, n = n_1(l), p \in P_l \setminus \{1\} \tag{3}$$

$$\text{arr}^n_{lp} - \text{arr}^n_{l(p-1)} \leq h^{max}_l \qquad l \in L, n = n_1(l), p \in P_l \setminus \{1\} \tag{4}$$

$$\text{dep}^n_{lp} = \text{arr}^n_{lp} + \text{dwt}^n_{lp} \qquad l \in L, n \in TN_l, p \in P_l \tag{5}$$

$$\text{arr}^n_{lp} = \text{dep}^{n_{prev}}_{lp} + ttd^{n_{prev}-n}_l \qquad l \in L, n \in TN_l \setminus \{n_1(l)\}, p \in P_l \tag{6}$$

Constraints (2) adjust the arrival time of the first trip of each line at its first transfer node to less than the maximum allowable value of the line's headway. Agencies can choose this upper limit based on their service standards. We define constraints (3) and (4) to ensure the given headway ranges advised by agencies, minimum and maximum allowable headways, are applied for the first transfer node of each line. However, headway regularity cannot be guaranteed for the following nodes since it depends on dwell times, and if we force headway regularity, the model may end up infeasible. We discuss further details about headway regularity and provide some soft constraints on them at the end of this section.

Constraints (5) calculate the departure time of each bus from each transfer node. A bus departure time equals its arrival time at the node plus total dwell time. A bus dwell time depends on the number of alighting and boarding passengers, whichever is larger. The total boarding time includes different service times depending on the number of successful boarding passengers considering bus capacity and assigned transfer buffer times. The required constraints for dwell time computation are discussed in the following. Note that if a bus is not a connecting bus, its dwell time only depends on the number of local boarding and alighting passengers. In constraints (6), the arrival time of each bus at a transfer node (except its first node) is determined based on its departure time from the previous node and the running time between the two consecutive transfer nodes.

$$(\text{arr}^n_{lp} + awt^n_{ll'}) - (\text{arr}^n_{l'q} + \text{dtb}^n_{l'q}) \leq (1 - \text{Y}^n_{lpl'q})M_1 \qquad (n, l, l') \in TP, q \in Q_{l'}, p \in P_l \tag{7}$$

$$(\text{arr}^n_{lp} + awt^n_{ll'}) - (\text{arr}^n_{l'q} + \text{dtb}^n_{l'q}) \geq (-\text{Y}^n_{lpl'q})M_2 + \epsilon \qquad (n, l, l') \in TP, q \in Q_{l'}, p \in P_l \tag{8}$$

$$(\text{arr}^n_{lp} + awt^n_{ll'}) - (\text{arr}^n_{l'q} + \text{serv1}^n_{l'q} + \text{serv2}^n_{l'q} + \text{tb}^n_{l'q}) \leq (1 - \text{YY}^n_{lpl'q})M_3 \qquad (n, l, l') \in TP, q \in Q_{l'}, p \in P_l \tag{9}$$

$$(\text{arr}^n_{lp} + awt^n_{ll'}) - (\text{arr}^n_{l'q} + \text{serv1}^n_{l'q} + \text{serv2}^n_{l'q} + \text{tb}^n_{l'q}) \geq (\text{YY}^n_{lpl'q} - 1)M_4 \qquad (n, l, l') \in TP, q \in Q_{l'}, p \in P_l \tag{10}$$

$$\text{ZY}^n_{lpl'q} = \text{Y}^n_{lpl'q} \qquad (n, l, l') \in TP, q = 1, p \in P_l \tag{11}$$

$$\text{ZY}^n_{lpl'q} = \text{Y}^n_{lpl'q} - \text{Y}^n_{lpl'(q-1)} \qquad (n, l, l') \in TP, q \in Q_{l'} \setminus \{1\}, p \in P_l \tag{12}$$

$$\text{SY}^n_{lpl'} = \sum_{q \in Q_{l'}} (\text{ZY}^n_{lpl'q}) - \sum_{q \in Q^-_{l'}} (\text{YY}^n_{lpl'q}) \qquad (n, l, l') \in TP, p \in P_l \tag{13}$$

$$\text{TY}^n_{lpl'q} \leq \text{SY}^n_{lpl'} \qquad (n, l, l') \in TP, q \in Q_{l'}, p \in P_l \tag{14}$$

$$\text{TY}^n_{lpl'q} \geq (\text{SY}^n_{lpl'} - 1) + \text{ZY}^n_{lpl'q} \qquad (n, l, l') \in TP, q \in Q_{l'}, p \in P_l \tag{15}$$

$$\sum_{q \in Q_{l'}} (\text{TY}^n_{lpl'q} + \text{YY}^n_{lpl'q}) + \text{TY}^n_{lpl'q^*} = 1 \qquad (n, l, l') \in TP, p \in P_l \tag{16}$$

In constraints (7) and (8), if transferring passengers of the feeder bus p of line l arrive at the connecting bus stop location later than the arrival time plus the DTB of the connecting bus q of line l', $\texttt{dtb}^n_{l'q}$, then those passengers cannot make a successful transfer of Types 1 or 2, i.e., $\text{Y}^n_{lpl'q}$ equals zero; otherwise $\text{Y}^n_{lpl'q}$ equals one. Nevertheless, if $\text{Y}^n_{lpl'q}$ equals one, it does not necessarily mean that transfer Types 1 or 2 can occur. In other words, constraints (7) and (8) can only recognize the non-eligible and eligible connecting buses associated with successful transfer Types 1 and 2, but not the exact assigned connecting bus for a transfer. Note that if passengers' arrival time from the feeder bus p of line l is equal to arrival time plus the DTB of the connecting bus q of line l', then the left-hand sides are zero in both constraints (7) and (8). Since, in this case, passengers can make a successful transfer, we use ϵ in constraints (8) to ensure that $\mathtt{Y}^{n}_{lpl'q}$ equals one.

On the other hand, if $\mathtt{YY}^{n}_{lpl'q}$ in constraints (9) and (10) equals one, it means that bus q of line l' is the assigned connecting bus, as successful transfer Type 3, for transferring passengers of vehicle p of line l. It is worth mentioning that, regardless of the differences between the arrival of the feeder and connecting buses, $\mathtt{YY}^{n}_{lpl'q}$ equals zero can always satisfy constraints (9) and (10). However, if a successful transfer occurs through Type 3, constraints (9) and (10) can determine the assigned connecting bus by assigning $\mathtt{YY}^{n}_{lpl'q}$ to one. In this case, the lefthand side equals zero, meaning that the arrival time of transferring passengers equals exactly the arrival time of the connecting bus plus the first and second service times and the assigned transfer buffer time. This decision will be made with the help of other constraints that are explained in the following, constraints (11)-(16) and (27)-(32), where the goal is assigning at least one connecting bus to a feeder bus with the minimum possible transfer waiting time. We note that $\mathtt{tb}^{n}_{l'q}$ can have a negative value, meaning that the transferring passengers of vehicle p of line l arrive during $\mathtt{serv2}^{n}_{l'q}$ of vehicle q of line l'. In this case, the connecting vehicle is not being held, and no penalty must be considered. Therefore, another variable, ptb, is defined as the positive transfer buffer, which equals TB if there is a positive transfer buffer time, i.e., holding time, assigned to a bus; otherwise, it equals zero. We also use the ptb in calculating a bus dwell time.

Now, let us explain when exactly successful transfer Types 1 and 2 occur. When $\mathtt{Y}^{n}_{lpl'q}$ equals one in both constraints (7) and (8), it means that bus q of line l' is only an eligible connection for transferring passengers of bus p of line l at node n. Hence, the model cannot yet decide which type of successful transfer would occur and it needs to check whether the previous connecting bus, $(q-1)$, is the actual connecting bus through successful transfer Type 3 or not. To help the model distinguish transfer Types 1 and 2 from transfer Type 3, three other indicator variables, namely, ZY, TY, and SY, are defined. In constraints (11) and (12), the ZY variable finds the first eligible bus for Types 1 and 2, i.e., the first connecting bus with $\mathtt{Y}^{n}_{lpl'q}$ equals one. Then, in order to assign precisely one type of successful transfer, i.e., one connecting bus to each feeder bus, constraints (13)-(16) are defined.

In constraints (13), if SY is zero, two possible conditions could happen: (I) both summations equal zero, (II) both summations equal one. Condition (I) means that all ZY and YY variables in both summations are zero, so there are two ways to have a successful transfer: either to have a successful transfer Type 3 with the last bus of line l', or to have a successful transfer via the first bus from the next horizon (based on constraints (28)). If condition (II) is true, successful transfer Type 3 is the actual one, since when any $\mathtt{YY}^{n}_{lpl'q}$ equals one, the successful transfer happens using TB. So, the first summation in constraints 2 only shows the eligibility of the following bus, $q+1$ that comes after bus q for transfer Types 1 and 2, i.e., $\mathtt{ZY}^{n}_{lpl'(q+1)}$ and $\mathtt{YY}^{n}_{lpl'(q+1)}$ equal one. Note that in a case that all ZY variables are zero, and the last trip with the help of TB would be the actual successful transfer, SY would equal negative one if we add all the q buses; thus, the second summation excludes the last trip of the connecting line.

Variable TY is defined to confirm whether a successful transfer would occur through Types 1 or 2. As explained above, if SY equals zero, then Types 1 and 2 of successful transfer cannot happen, thus TY should be zero, which is ensured by (14). On the other hand, if SY equals one, it means that transfer Type 1 or 2 can happen, so TY equals one for the correct connecting bus q through constraints (14) and (15). Constraints (16) ensure that each feeder bus has an assigned connecting bus through any type of successful transfers or the first trip of the next horizon.

$$\mathrm{dtb}^{n}_{l'q} \leq \mathrm{serv1}^{n}_{l'q} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{17}$$

$$(-\mathrm{tb}^{n}_{l'q}) \leq \mathrm{serv2}^{n}_{l'q} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{18}$$

$$\mathrm{tb}^{n}_{l'q} \leq (\mathrm{Ib}^{n}_{l'q})M_5 \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{19}$$

$$\mathrm{tb}^{n}_{l'q} \geq (\mathrm{Ib}^{n}_{l'q} - 1)M_5 + \epsilon \qquad (n \in N, l' \in CL_n, q \in Q_{l'} \tag{20}$$

$$(\mathrm{ptb}^{n}_{l'q}) \leq \mathrm{tb}^{n}_{l'q} + (1 - \mathrm{Ib}^{n}_{l'q})M_6 \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{21}$$

$$(\mathrm{ptb}^{n}_{l'q}) \geq \mathrm{tb}^{n}_{l'q} + (\mathrm{Ib}^{n}_{l'q} - 1)M_6 \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{22}$$

$$(\mathrm{ptb}^{n}_{l'q}) \leq (\mathrm{Ib}^{n}_{l'q})M_6 \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{23}$$

$$(\mathrm{ptb}^{n}_{l'q}) \geq (-\mathrm{Ib}^{n}_{l'q})M_6 \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{24}$$

Constraints (17) ensure that the value of DTB does not exceed the service time of local passengers and Type 1 successfully transferred passengers. Similarly, Constraints (18) bound the negative value of TB by the service time of

Type 2 successfully transferred passengers. Constraints (19) and (20) identify whether any positive TB is assigned to a bus. If we have negative or zero TB assigned to a bus, $\mathtt{Ib}^{n}_{l'q}$ equals zero. Accordingly, constraints (21)-(24) determine the value of $\mathtt{ptb}^{n}_{l'q}$, equal to the transfer buffer if positive, zero otherwise, which is used in dwell time calculation. If $\mathtt{Ib}^{n}_{l'q}$ equals one, we need to consider the penalty for in-vehicle passengers who are being held due to TB.

We also want to emphasize that two other restrictions regarding the value of TB can be added as constraints, which agencies could set based on their service standards, if any: TB at a transfer node for each bus should be less than a specific value, constraints (25), and the total value of TB for a bus along all its transfer nodes should be less than a particular value, constraints (26). These limits can be chosen based on a line's headway or a trip's layover time.

$$\text{ptb}^{n}_{l'q} \le Mtb^{n}_{l'} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{25}$$

$$\sum_{n \in TN_l} (\text{ptb}^{n}_{l'q}) \le Ttb^{max}_{l'} \qquad l' \in L, q \in Q_{l'} \tag{26}$$

Constraints (27) and (28) determine the waiting time of transferring passengers, which is the variable included in the objective function with its associated transferring demand. If successful transfer Types 2 or 3 occur, the value of `NTwait` equals zero since the arrival of transferring passengers would be later than the arrival of their connection. Note that to ensure the feasibility of constraints (16), we also consider a possible successful transfer to a connecting trip from the next planning horizon, q^*. However, we define a high cost for this connection to force the model to assign connecting buses from the current planning horizon. Constraints (28) are defined for this aim. In order to differentiate between successful transfer Types 1 and 2, in which `TY` equals one for both, we also define constraints (29)-(32). We need to separate Types 1 and 2 to determine the number of passengers in each type of successful transfer to calculate their service time for boarding. If transferring passengers arrive before the arrival of their assigned connecting bus, the value of transferring waiting time, `NTwait`, is more than zero and is determined by constraints (27). Accordingly, no matter we have `dtb` or not, `Twait`, as computed in constraints (29) and (30) is also positive for these passengers. If the difference between `Twait` and `NTwait` is more than zero, it can be only less than or equal to the value of `dtb`. If the difference is exactly equal to `dtb`, it means that passengers would arrive before the arrival of their assigned connecting bus, i.e., successful transfer Type 1. In this case, `Iwait` equals 1 in constraints (31) and (32). On the other hand, while `TY` equals 1, the difference cannot be less than `dtb` or zero unless `NTwait` equals zero. In this condition, passengers would arrive during or at the end of `dtb`. Thus, the left-hand sides in constraints (31) and (32) are positive leading to `Iwait` to be zero. Note that we have to ensure that `Iwait` equals one if and only if successful transfer Type 1 can occur. Therefore, we have to enforce `Iwait` equals zero in other possible cases, including no successful transfer when `TY` equals zero and `NTwait` is also zero. To enforce `Iwait` equals zero, we define constraints (29) and (30) in a way to provide a large negative number in case of `TY` equals zero to force constraints (31) and (32) to have `Iwait` equals zero. Otherwise, if `Twait` is a small value or zero, `Iwait` would be one when actually no transfer would occur.

$$\text{arr}^{n}_{l'q} - (\text{arr}^{n}_{lp} + awt^{n}_{ll'}) + (\text{TY}^{n}_{lpl'q} - 1)M_7 \le \text{NTwait}^{n}_{lpl'q} \qquad (n,l,l') \in TP, q \in Q_{l'}, p \in P_l \tag{27}$$

$$\text{NTwait}^{n}_{lpl'q^*} \ge (\text{TY}^{n}_{lpl'q^*} - 1)M_8 + M_9 \qquad (n,l,l') \in TP, p \in P_l \tag{28}$$

$$(\text{arr}^{n}_{l'q} + \text{dtb}^{n}_{l'q}) - (\text{arr}^{n}_{lp} + awt^{n}_{ll'}) + (\text{TY}^{n}_{lpl'q} - 1)M_{10} \le \text{Twait}^{n}_{lpl'q} \qquad (n,l,l') \in TP, q \in Q_{l'}, p \in P_l \tag{29}$$

$$(\text{arr}^{n}_{l'q} + \text{dtb}^{n}_{l'q}) - (\text{arr}^{n}_{lp} + awt^{n}_{ll'}) + (\text{TY}^{n}_{lpl'q} - 1)M_{11} \ge \text{Twait}^{n}_{lpl'q} \qquad (n,l,l') \in TP, q \in Q_{l'}, p \in P_l \tag{30}$$

$$\text{dtb}^{n}_{l'q} - (\text{Twait}^{n}_{lpl'q} - \text{NTwait}^{n}_{lpl'q}) \le (1 - \text{Iwait}^{n}_{lpl'q})M_{12} \qquad (n,l,l') \in TP, p \in P_l, q \in Q_{l'} \tag{31}$$

$$\text{dtb}^{n}_{l'q} - (\text{Twait}^{n}_{lpl'q} - \text{NTwait}^{n}_{lpl'q}) \ge (-\text{Iwait}^{n}_{lpl'q})M_{13} + \epsilon \qquad (n,l,l') \in TP, p \in P_l, q \in Q_{l'} \tag{32}$$

As discussed above, we consider the connecting bus arrival times instead of departure times for defining different types of successful transfers in our formulation. The reason is that the actual departure time occurs when all passengers are on-board, i.e., need to be determined, considering the required service time to board. Therefore, a bus dwell time is calculated based on each bus' number of boarding passengers considering the bus' available capacity through the constraints explained next. Although the model attempts to assign buses with spare capacity while minimizing transfer waiting times, it is still possible that the optimal timetables result in leftover passengers. Nevertheless, the benefit of this formulation is to recognize if there is a considerable capacity limitation based on the current resources, i.e., number of trips and bus size.

Constraints (33) determine the number of in-vehicle passengers when a bus arrives at a transfer node, ivd. This value for a bus at its first node is a given number estimated from historical data. However, for the subsequent transfer nodes, ivd depends on the number of alighting, $ad_{l'q}^{n_{prev}}$, and successful boarding passengers at the previous transfer node, $\mathtt{Tbd}_{l'q}^{n_{prev}}$, plus the total net value of boarding and alighting at the intermediate stops between the two transfer nodes, $sp_{l'q}^{n_{prev}-n}$. The number of total boarding, $\mathtt{Tbd}_{l'q}^{n_{prev}}$, depends on the bus capacity which is determined separately as explained later.

$$\text{ivd}_{l'q}^{n} = \text{ivd}_{l'q}^{n_{prev}} - ad_{l'q}^{n_{prev}} + \text{Tbd}_{l'q}^{n_{prev}} + sp_{l'q}^{n_{prev}-n} \qquad n \in N \setminus \{n_1(l')\}, l' \in CL_n, q \in Q_{l'} \quad (33)$$

We break down the calculation of the exact dwell time, dwt, into three parts. The first part, serv1, depends on the number of the first group of boarding passengers, i.e., local passengers, ld, and successfully transferred passengers who arrive at their connecting bus location before or exactly at the same time as the arrival time of their connection, i.e., successful transfer Type 1. As explained above, when Iwait equals one, we can determine the number of successful transfer Type 1. Constraints (33) determine the first group of generated demand for boarding.

$$\text{GBD1}_{l'q}^{n} = \sum_{l \in L:(n,l,l') \in TP} \sum_{p \in P_l} \text{Iwait}_{lpl'q}^{n} td_{lpl'}^{n} + ld_{l'q}^{n} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (34)$$

We also consider bus capacity in our model, so we need to determine the proportion of the generated demand in constraints (34) who can actually board their bus. Constraints (35) compute the existing capacity of a connecting bus q of line l' which equals the difference of the nominal vehicle capacity minus the number of in-vehicle passengers, when the bus arrives at a node, plus the number of passengers who alight from the bus at that node. Constraints (36) and (37) examine whether there is enough space for the first generated demand, GBD1, and the remaining first leftover passengers from the previous bus, $(q - 1)$. If CI1 equals zero, it means that the existing capacity of the bus is more than needed so all the passengers can board the bus. Constraints (38)-(41) determine the number of passengers who are able to get on bus q of line l' at node n, as the first boarding group. If the existing capacity is less than the demand, CI1 equals one, then the value of total boarding is equal to the existing capacity. Accordingly, the number of leftover passengers will be determined by constraints (66)-(73). Also, the first service time, serv1, based on the number of boarding passengers, Tbd1, is calculated by constrains (42). Note that when q equals one the value of $\mathtt{flp}_{l'(q-1)}^{n}$ is given and considered to be zero.

$$\text{EC}_{l'q}^{n} = vcn_{l'q} - \text{ivd}_{l'q}^{n} + ad_{l'q}^{n} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (35)$$

$$\text{EC}_{l'q}^{n} - (\text{GBD1}_{l'q}^{n} + \text{flp}_{l'(q-1)}^{n} - \text{slp}_{l'(q-2)}^{n}) \leq (1 - \text{CI1}_{l'q}^{n})M_{14} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (36)$$

$$\text{EC}_{l'q}^{n} - (\text{GBD1}_{l'q}^{n} + \text{flp}_{l'(q-1)}^{n} - \text{slp}_{l'(q-2)}^{n}) \geq (-\text{CI1}_{l'q}^{n})M_{15} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (37)$$

$$\text{Tbd1}_{l'q}^{n} \geq (\text{EC}_{l'q}^{n}) + (\text{CI1}_{l'q}^{n} - 1)M_{16} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (38)$$

$$\text{Tbd1}_{l'q}^{n} \leq \text{EC}_{l'q}^{n} + (1 - \text{CI1}_{l'q}^{n})M_{18} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (39)$$

$$\text{Tbd1}_{l'q}^{n} \geq (-\text{CI1}_{l'q}^{n})M_{17} + (\text{GBD1}_{l'q}^{n} + \text{flp}_{l'(q-1)}^{n} - \text{slp}_{l'(q-2)}^{n}) \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (40)$$

$$\text{Tbd1}_{l'q}^{n} \leq (\text{CI1}_{l'q}^{n})M_{19} + (\text{GBD1}_{l'q}^{n} + \text{flp}_{l'(q-1)}^{n} - \text{slp}_{l'(q-2)}^{n}) \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (41)$$

$$\text{serv1}_{l'q}^{n} = \text{Tbd1}_{l'q}^{n} b^{t} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (42)$$

The next step in calculating a bus dwell time is to determine the number of passengers who get to their connecting bus through successful transfer Type 2. As explained above, if TY equals one either transfer Type 1 or 2 has occurred and if TY equals one and Iwait also equals one, it is transfer Type 1. So if TY equals one and Iwait equals zero, it means that passengers make their transfer via Type 2. This demand, GBD2, is determined in constraints (43). Similar to GBD1, we need to check whether the bus has enough capacity for GBD2 or not. Constraints (44)-(47) are defined for this aim. Accordingly, the required dwell time, serv2, is calculated in constraints (50).

$$\text{GBD2}_{l'q}^{n} = \sum_{l \in L:(n,l,l') \in TP} \sum_{p \in P_l} (\text{TY}_{lpl'q}^{n} - \text{Iwait}_{lpl'q}^{n}) td_{lpl'}^{n} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (43)$$

$$(\text{EC}_{l'q}^{n} - \text{Tbd1}_{l'q}^{n}) - \text{GBD2}_{l'q}^{n} \leq (1 - \text{CI2}_{l'q}^{n})M_{20} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \quad (44)$$

$$(\mathrm{EC}^n_{l'q} - \mathrm{Tbd1}^n_{l'q}) - \mathrm{GBD2}^n_{l'q} \geq (-\mathrm{CI2}^n_{l'q})M_{21} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{45}$$

$$\mathrm{Tbd2}^n_{l'q} \geq (-\mathrm{CI2}^n_{l'q})M_{23} + \mathrm{GBD2}^n_{l'q} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{46}$$

$$\mathrm{Tbd2}^n_{l'q} \leq (\mathrm{CI2}^n_{l'q})M_{25} + \mathrm{GBD2}^n_{l'q} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{47}$$

$$\mathrm{Tbd2}^n_{l'q} \geq (\mathrm{EC}^n_{l'q} - \mathrm{Tbd1}^n_{l'q}) + (\mathrm{CI2}^n_{l'q} - 1)M_{22} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{48}$$

$$\mathrm{Tbd2}^n_{l'q} \leq (\mathrm{EC}^n_{l'q} - \mathrm{Tbd1}^n_{l'q}) + (1 - \mathrm{CI2}^n_{l'q})M_{24} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{49}$$

$$\mathrm{serv2}^n_{l'q} = \mathrm{Tbd2}^n_{l'q} b^t \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{50}$$

The last part of the dwell time variable, `dwt`, depends on the assigned TB of a bus and the number of passengers who get to their connecting bus with the help of TB, i.e., successful transfer of Type 3. The model first determines the total number of passengers, successfully transferred and the local ones, generated for a connecting bus through constraints (51). Then based on the same logic we used above, constraints (52)-(58) check the bus capacity and the total number of passengers who are able to board, `Tbd`. The extra service time, `serv3`, for the passengers who get to their connecting bus with the help of TB is calculated through constraints (58). Note that in the case of having a greater number of alighting passengers than boarding passengers, the total dwell time depends on the former. Therefore, we need to check which one is the longest one and then assign the total dwell time to a bus. This comparison is done through constraints (59) and (60). Then constraints (61)-(64) determine a bus dwell time, `dwt`.

$$\mathrm{Ttd}^n_{l'q} = \sum_{l \in L:(n,l,l') \in TP} \sum_{p \in P_l} (\mathrm{TY}^n_{lpl'q} + \mathrm{YY}^n_{lpl'q}) td^n_{lpl'} + ld^n_{l'q} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{51}$$

$$\mathrm{EC}^n_{l'q} - (\mathrm{Ttd}^n_{l'q} + \mathrm{flp}^n_{l'(q-1)} - \mathrm{slp}^n_{l'(q-2)}) \leq (1 - \mathrm{CI}^n_{l'q})M_{26} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{52}$$

$$\mathrm{EC}^n_{l'q} - (\mathrm{Ttd}^n_{l'q} + \mathrm{flp}^n_{l'(q-1)} - \mathrm{slp}^n_{l'(q-2)}) \geq (-\mathrm{CI}^n_{l'q})M_{27} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{53}$$

$$\mathrm{Tbd}^n_{l'q} \leq \mathrm{EC}^n_{l'q} + (1 - \mathrm{CI}^n_{l'q})M_{30} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{54}$$

$$\mathrm{Tbd}^n_{l'q} \geq \mathrm{EC}^n_{l'q} + (\mathrm{CI}^n_{l'q} - 1)M_{28} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{55}$$

$$\mathrm{Tbd}^n_{l'q} \leq (\mathrm{CI}^n_{l'q})M_{31} + (\mathrm{Ttd}^n_{l'q} + \mathrm{flp}^n_{l'(q-1)} - \mathrm{slp}^n_{l'(q-2)}) \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{56}$$

$$\mathrm{Tbd}^n_{l'q} \geq (-\mathrm{CI}^n_{l'q})M_{29} + (\mathrm{Ttd}^n_{l'q} + \mathrm{flp}^n_{l'(q-1)} - \mathrm{slp}^n_{l'(q-2)}) \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{57}$$

$$\mathrm{serv3}^n_{l'q} = (\mathrm{Tbd}^n_{l'q} - \mathrm{Tbd2}^n_{l'q} - \mathrm{Tbd1}^n_{l'q})(b^t) \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{58}$$

$$(\mathrm{serv1}^n_{l'q} + \mathrm{serv2}^n_{l'q} + \mathrm{ptb}^n_{l'q} + \mathrm{serv3}^n_{l'q}) - (ad^n_{l'q})a^t \leq (1 - \mathrm{TI}^n_{l'q})M_{32} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{59}$$

$$(\mathrm{serv1}^n_{l'q} + \mathrm{serv2}^n_{l'q} + \mathrm{ptb}^n_{l'q} + \mathrm{serv3}^n_{l'q}) - (ad^n_{l'q})a^t \geq (-\mathrm{TI}^n_{l'q})M_{33} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{60}$$

$$\mathrm{dwt}^n_{l'q} \leq (\mathrm{serv1}^n_{l'q} + \mathrm{serv2}^n_{l'q} + \mathrm{ptb}^n_{l'q} + \mathrm{serv3}^n_{l'q}) + (1 - \mathrm{TI}^n_{l'q})M_{34} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{61}$$

$$\mathrm{dwt}^n_{l'q} \geq (\mathrm{serv1}^n_{l'q} + \mathrm{serv2}^n_{l'q} + \mathrm{ptb}^n_{l'q} + \mathrm{serv3}^n_{l'q}) + (\mathrm{TI}^n_{l'q} - 1)M_{36} \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{62}$$

$$\mathrm{dwt}^n_{l'q} \leq (\mathrm{TI}^n_{l'q})M_{35} + (ad^n_{l'q})a^t \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{63}$$

$$\mathrm{dwt}^n_{l'q} \geq (-\mathrm{TI}^n_{l'q})M_{37} + (ad^n_{l'q})a^t \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{64}$$

As mentioned above, we need to consider the penalty of extra stopping time, TB, for in-vehicle passengers. The passengers who would be penalized due to TB include through passengers, local passengers, and successfully transferred passengers via Types 1 and 2. Through passengers arrive from the previous node and do not alight from the bus at the current node. At the first node, the number of in-vehicle passengers is given. However, for the remaining transfer nodes of each line, based on the number of boarding and alighting passengers at each node, it needs to be determined and tracked node to node. Hence, the number of in-vehicle passengers is a decision variable, thus multiplying it by the exact value of the transfer time would lead to a non-linear objective. To prevent this, we only consider a fixed penalty value for the passengers who are penalized by TB and `serv3`. Moreover, we note that transfer buffer penalty should be considered when passengers would be held while there is no service required, either boarding or alighting. Therefore, if alighting time is more than boarding time, i.e., `TI` equals zero, in-vehicle passengers would not be penalized. In constraints (65), if both `ib` and `TI` equal one, then `ptb` shows the number of penalized passengers; otherwise, if either of ib or/and TI equals zero then ptb would be zero.

$$\text{pd}^n_{l'q} \geq (\text{ivd}^n_{l'q} - ad^n_{l'q} + \text{Tbd1}^n_{l'q} + \text{Tbd2}^n_{l'q}) + (\text{Ib}^n_{l'q} - 1)M_{38} + (\text{TI}^n_{l'q} - 1)M_{38} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{65}$$

If the existing capacity is less than the total generated demand, then the number of total boarding equals the existing capacity, and the number of first leftover passengers is determined by constraints (66)-(73). Constraints (74)-(77) determine the number of passengers who not only miss their first assigned connecting bus due to lack of space but also they cannot get on the next arriving bus. The penalty for this group of passengers is higher compared to the first group of leftover passengers. It is assumed that they will exit the system, i.e., shift their mode. Note that $\texttt{flp}^n_{l'(q-1)}$, $\texttt{slp}^n_{l'(q-1)}$ and $\text{slp}^n_{l'(q-2)}$ for $\texttt{q} = 1$ are given and fixed to zero. Furthermore, unlike the approach applied in our previous work (Ansarilari et al., 2018), in this study, we have to explicitly determine the number of leftover passengers. In other words, we cannot rely on the objective function to choose the value of zero for slp or flp if the capacity is sufficient. The reason is that in some cases, the model might assign leftover passengers in favor of bus dwell times by inaccurate values of leftover and boarding passengers.

$$\text{Ttd}^n_{l'q} - (\text{EC}^n_{l'q} - \text{flp}^n_{l'(q-1)} + \text{slp}^n_{l'(q-2)}) \leq (1 - \text{FL}^n_{l'q})M_{39} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{66}$$

$$\text{Ttd}^n_{l'q} - (\text{EC}^n_{l'q} - \text{flp}^n_{l'(q-1)} + \text{slp}^n_{l'(q-2)}) \geq (-\text{FL}^n_{l'q})M_{40} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{67}$$

$$\text{flp}^n_{l'q} \leq (1 - \text{FL}^n_{l'q})M_{41} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{68}$$

$$\text{flp}^n_{l'q} \geq (\text{FL}^n_{l'q} - 1)M_{42} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{69}$$

$$\text{flp}^n_{l'q} \leq \text{Ttd}^n_{l'q} - (\text{EC}^n_{l'q} - \text{flp}^n_{l'(q-1)} + \text{slp}^n_{l'(q-2)}) + (\text{FL}^n_{l'q})M_{43} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{70}$$

$$\text{flp}^n_{l'q} \geq \text{Ttd}^n_{l'q} - (\text{EC}^n_{l'q} - \text{flp}^n_{l'(q-1)} + \text{slp}^n_{l'(q-2)}) + (-\text{FL}^n_{l'q})M_{44} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{71}$$

$$\text{flp}^n_{l'(q-1)} - \text{EC}^n_{l'q} \leq (1 - \text{SL}^n_{l'q})M_{45} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{72}$$

$$\text{flp}^n_{l'(q-1)} - \text{EC}^n_{l'q} \geq (-\text{SL}^n_{l'q})M_{46} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{73}$$

$$\text{slp}^n_{l'(q-1)} \leq (1 - \text{SL}^n_{l'q})M_{47} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{74}$$

$$\text{slp}^n_{l'(q-1)} \geq (\text{SL}^n_{l'q} - 1)M_{48} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{75}$$

$$\text{slp}^n_{l'(q-1)} \leq \text{flp}^n_{l'(q-1)} - \text{EC}^n_{l'q} + (\text{SL}^n_{l'q})M_{49} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{76}$$

$$\text{slp}^n_{l'(q-1)} \geq \text{flp}^n_{l'(q-1)} - \text{EC}^n_{l'q} + (-\text{SL}^n_{l'q})M_{50} \quad n \in N, l' \in CL_n, q \in Q_{l'} \tag{77}$$

Lastly, we discuss some potential additions to the presented baseline model (1)-(77). As mentioned above, we restricted the bus arrivals at the first transfer nodes of each line to occur between the minimum and maximum allowable line's headway. However, headway regularity cannot be guaranteed at every node. This is because we are not only determining the exact dwell time of buses but also adding extra transfer buffer times. Therefore, we define soft constraints for headway regularity between the arrival of two consecutive buses of a line at each transfer node, except the first one, constraints (78) and (79). If a headway range is violated, a penalty would be added to the objective function. Note that $\texttt{arr}^n_{l'(q-1)}$ for $\texttt{q} = 1$ is fixed to zero. It is worth mentioning that in previous timetabling studies at the scheduling stage, headway regularity was also only considered for departures from terminals, and the arrivals/departures times at other transfer nodes were not forced to be in the range of acceptable headway. On the other hand, headway regularity has been mostly maintained in the operational stage at control stops by real-time strategies. Nevertheless, in our study, by considering a detailed approach for the dwell time determination, the arrival and departure times are designed more accurately, providing an opportunity to consider headway regularity at the scheduling stage.

$$h^{min}_{l'} - (\text{arr}^n_{l'q} - \text{arr}^n_{l'(q-1)}) \leq hreg^n_{l'q} \quad n \in N \setminus \{n_1(l')\}, l' \in CL_n, q \in Q_{l'} \tag{78}$$

$$(\text{arr}^n_{l'q} - \text{arr}^n_{l'(q-1)}) - h^{max}_{l'} \leq hreg^n_{l'q} \quad n \in N \setminus \{n_1(l')\}, l' \in CL_n, q \in Q_{l'} \tag{79}$$

As discussed above, the model considers the penalty of transfer buffer time for the on-board passengers. Thus, we only consider local passengers coming from the streets as the main group of passengers affected by headway irregularity. By multiplying the headway deviations by the number of local passengers, expression (80), incorporated into the objective function, we can explicitly take into consideration the passengers who are affected by headway irregularity. This will be discussed more in Section 7.3, based on a numerical experiment.

$$\sum_{n\in N}\sum_{l'\in CL_n}\sum_{q\in P_{l'}} ld^n_{l'q} hreg^n_{l'q} c^h \tag{80}$$

6. Solution Method: Model Reformulation Based on Lagrangian Relaxation Approach

Although we have an extensive model with a large number of constraints and variables, it has a weakly coupled structure, where coupling happens via constraints (6) and (33), which facilitates breaking it down into smaller node-based optimization problems. As presented in Section 5, all the constraints in our model are formulated for each node individually, except constraints (6) and (33). Constraints (6) connect each bus's arrival and departure times from a transfer node to the next one considering the running time between nodes. Similarly, constraints (33) track the in-vehicle demand of buses along a line's sequence of transfer nodes. Inspired by the special structure in our model, as a solution method, we apply an iterative procedure, where at each iteration, a valid lower bound on the optimal value of our MIP model as well as a feasible solution to the model (thus an upper bound on the optimal value) are generated. We continue the iterations until we reach an acceptable gap between the bounds. Subsequently, we either consider the feasible solution corresponding to the best upper-bound value as the final output, or we use that feasible solution as a warm start to the original MIP model via a commercial MIP solver, such as GUROBI, and solve it to optimality. We employ the Lagrangian relaxation method and a heuristic optimization model to apply this procedure, which are explained in detail next.

The Lagrangian relaxation (LR) approach is a well-proven method to efficiently generate high-quality solutions by relaxing "problematic" constraints of a model, thus reducing the solution time considerably (Vera, Weintraub, Koenig, Bravo, Guignard and Barahona, 2003). The candidate constraints for relaxation can be chosen based on the structure of the model or through an initial analysis to identify the constraints that cause the model to suffer from a computational burden of high proportions. In the LR approach, the relaxation of the chosen constraints is integrated into the model's objective function as penalization elements for any unit of their violation. The penalization component is the product of the constraints' violation magnitude and specified coefficients called Lagrange multipliers. A detailed discussion and explanation regarding the Lagrangian relaxation approach can be found in (Vera et al., 2003).

In our model, by relaxing constraints (6) and (33) and incorporating the relaxation penalization components into the objective function, the model transforms into a fully node-based formulation thus can be decomposed into separate node-specific optimization problems, called subproblems. Thus, rather than solving the original model, i.e., optimize all the nodes simultaneously, each subproblem is solved individually without directly considering its effects on the other nodes and the effects of other nodes on itself. The only indirect connection between the nodes is now through Lagrangian multipliers.

Although we can relax both constraints (6) and (33), we alternatively remove constraints (33) from the original model before applying the Lagrangian relaxation. This strategy simplifies our Lagrangian-based solution method considerably while not impacting our formulation logic. The main reason is that the removal of constraints (33), i.e., not tracking the in-vehicle number of passengers, leads to adding the setting of unlimited capacity for boarding passengers. Hence, the relaxed model does not need to compare the generated demand variables, `GBD1`, `GBD2`, and `GBD3`, with the existing capacity of a bus, `EC`, to determine the successful boarding passengers, `Tbd1`, `Tbd2`, `Tb3`. Thus, the associated capacity-related constraints, (35)-(41), (44)-(49), and (52)-(57), and decision variables, `CI`, `CI2`, `CI3`, `EC`, `Tbd1`, `Tbd2`, and `Tbd3`, can be removed from the original model. This reduces the number of variables and constraints, resulting in a smaller feasible solution search space. Moreover, by removing constraints (33) instead of dualizing them through the LR approach (i.e., incorporating into the objective function), fewer Lagrangian multipliers need to be defined and optimized.

On the other hand, constraints (6) are essential to our formulation. In fact, they are crucial for generating timetables and line trajectories. Therefore, we relax them and consider their violations in the objective function using the LR approach. We dualize constraints (6) using the set of Lagrangian multipliers, λ^n_{lp}. Our obtained relaxed model, which we refer to as the `LR` model, is as follows:

$$LR(\lambda) = \min \sum_{(n,l,l')\in TP}\sum_{p\in P_l}\sum_{q\in Q^*_{l'}} td^n_{lpl'}(\text{NTwait}^n_{lpl'q})c^{tw}+ \tag{81a}$$

Table 4
Notation: Solution method

Sets:	
L_a^n	Set of lines at node n which will pass another transfer node in their own sequence after (a) node $n \in N$; $l \in L_a^n$
L_b^n	Set of lines at node n which have passed a transfer node in their sequence before (b) node $n \in N$; $l \in L_b^n$
Variables:	
SV^{nk}	Optimal objective function value of a subproblem for node n in iteration k; $n \in N$
λ_{lp}^{nk}	Lagrange mulitiplier for bus p of line l at node n in iteration k; $l \in L_b^n$, $p \in P_l$
sg_{lp}^{nk}	The elements of a subproblem's subgradient vector for bus p of line l at node n in iteration k; $l \in L_b^n$, $p \in P_l$
$\text{ONTwait}_{lpl'q}^{nk}$	Waiting times of each subproblem in iteration k; $(n,l,l') \in TP$, $p \in P_l$, $q \in Q_{l'}^*$
$\text{UNTwait}_{lpl'q}^{nk}$	Waiting times of the upper bound model in iteration k; $(n,l,l') \in TP$, $p \in P_l$, $q \in Q_{l'}^*$
$\text{DNTwait}_{lpl'q}^{nk}$	The difference between $\text{ONTwait}_{lpl'q}^{nk}$ and $\text{UNTwait}_{lpl'q}^{nk}$ in iteration k; $(n,l,l') \in TP$, $p \in P_l$, $q \in Q_{l'}^*$
$\text{ADNTwait}_{lpl'q}^{nk}$	The absolute value of $\text{DNTwait}_{lpl'q}^{nk}$ in iteration k; $(n,l,l') \in TP$, $p \in P_l$, $q \in Q_{l'}^*$
LB^k	The obtained lower bound in iteration k
LB	Best obtained lower bound
UB^k	The obtained upper bound in iteration k
UB	Best obtained upper bound
G^k	The gap between UB and LB at the end of iteration k
Parameters:	
θ	The stopping preferred gap between UB and LB

$$\sum_{(n,l')\in CL_n} \sum_{q\in Q_{l'}} \text{pd}_{l'q}^n (Mtb_{l'}^n) c^{vt} + \tag{81b}$$

$$\sum_{n\in N} \sum_{l\in L_b^n} \sum_{p\in P_l} \lambda_{lp}^n (\text{arr}_{lp}^n - ttd_l^{n_{prev}-n}) + \tag{81c}$$

$$\sum_{n\in N} \sum_{l\in L_a^n} \sum_{p\in P_l} \lambda_{lp}^{n_{next}} (-\text{dep}_{lp}^n) \tag{81d}$$

s.t. (2) − (32), (34), (43), (51), (59) − (65)

$$\text{serv1}_{l'q}^n = \text{GBD1}_{l'q}^n b^t \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{82}$$

$$\text{serv2}_{l'q}^n = \text{GBD2}_{l'q}^n b^t \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{83}$$

$$\text{serv3}_{l'q}^n = (\text{Ttd}_{l'q}^n - \text{GBD2}_{l'q}^n - \text{GBD1}_{l'q}^n) b^t \qquad n \in N, l' \in CL_n, q \in Q_{l'} \tag{84}$$

The (81a) and (81b) expressions, similar to (1a) and (1b), represent the total passengers' transfer waiting time and the penalty of transfer buffer times for the in-vehicle passengers, respectively. However, in determining the number of passengers experiencing the buffer time penalty, `pb`, we substitute the number of passengers when a bus arrives at a transfer node, `ivd`, in constraint (65) with a given value. This is similar to the original MIP model when the `ivd` is also given for the first node of each line's sequence. Moreover, due to the relaxation of the buses capacity, the generated demand variables, `GBD1`, `GBD2`, and `GBD3`, are also equal to the successful boarding passengers. Therefore, constraints (42), (50), (65) of the original model, for determining passengers' boarding times, are substituted with constraints (82)-(84). The (81c) and (81d) are defined as the relaxation of constraints (6). Since we want the relaxation to also decompose the original model into node-based optimization subproblems, we have to reformulate constraints (6) into a node-based structure as well. Constraints (6) are valid when more than one transfer node is in a line's sequence.

To incorporate their relaxed version into the objective function, we define Lagrangian multipliers based on the arrival node of each bus. In other words, λ^n_{lp} is defined for node n when constraints (6) show the relationship between arrival time at node n based on the trip's departure time from n_{prev}, the transfer node prior to node n. Thus, we have to ensure that the departure time variables are multiplied by the Lagrangian multipliers pertaining to the next node, n_{next}, as shown in (81d). Note that all the other constraints of the LR model are identical to the original model.

As discussed above, a subproblem is solved for each node separately, and the summation of the objective function values of all subproblems, which we denote by SV, is the value of the objective function of the LR model. This value constitutes a valid lower bound for on the optimal value of the original model. However, the bound quality depends on the values of Lagrangian multipliers which we need to optimize. There are common approaches such as subgradient and cutting-plane algorithms to determine Lagrange multipliers providing the best (largest) possible lower bound. After examining the two approaches via some preliminary experiments for our model, we decided to employ the subgradient method for our computational study.

The solution of the LR model generates a timetable for each node-based subproblem without considering the connections between the nodes. As such, the obtained solution may not be feasible to the original model, which is usually the case. In our problem, if we do not end up with a feasible solution after solving the LR model, a bus departure time from a transfer node and its arrival time at the following transfer node would not be connected, or/and there would be some bus capacity violations in the system. For example, the former infeasibility happens when a node's optimal subproblem solution suggests an arrival time for a bus which cannot be applied because of conflicts with the departure time of the same bus from the previous node's individual optimal subproblem solution. However, this infeasible solution can help us produce an actual feasible solution. Thus, we devise another heuristic procedure to transform such an infeasible solution into a feasible one, thus to also obtain an upper bound on the optimal value of our original MIP model. Note that the obtained feasible solution is not necessarily optimal to the original model.

6.1. The Upper-bounding Model

We develop another optimization model which takes a solution of the relaxed model, which is not feasible to the original model (i.e., the solution consists of disconnected bus trajectories and timetables), as an input and uses it as guidance to construct a feasible solution for the original MIP model. This upper-bounding model, which we refer to as the UM model, is as follows:

$$UM = \min \sum_{(n,l,l')\in TP} \sum_{p\in P_l} \sum_{q\in Q^*_{l'}} td^n_{lpl'}(\text{NTwait}^n_{lpl'q})c^{tw} + \tag{85a}$$

$$\sum_{(n,l')\in CL_n} \sum_{q\in P_{l'}} \text{pd}^n_{l'q}(Mtb^n_{l'})c^{vt} + \tag{85b}$$

$$\sum_{(n,l')\in CL_n} \sum_{q\in P_{l'}} \text{flp}^n_{l'q}(h^{max}_l)c^{fl} + \tag{85c}$$

$$\sum_{(n,l')\in CL_n} \sum_{q\in P_{l'}} \text{slp}^n_{l'q}(2h^{max}_l)c^{sl} + \tag{85d}$$

$$\sum_{(n,l,l')\in TP} \sum_{p\in P_l} \sum_{q\in Q^*_{l'}} td^n_{lpl'}(\text{ADNTwait}^n_{lpl'q})c^{tw} \tag{85e}$$

s.t. (2) – (77)

$$\text{NTwait}^n_{lpl'q} + \text{DNTwait}^n_{lpl'q} = \text{ONTwait}^n_{lpl'q} \qquad (n,l,l') \in TP, p \in P_l, q \in Q^*_{l'} \tag{86}$$

$$\text{ADNTwait}^n_{lpl'q} \geq \text{DNTwait}^n_{lpl'q} \qquad (n,l,l') \in TP, p \in P_l, q \in Q^*_{l'} \tag{87}$$

$$\text{ADNTwait}^n_{lpl'q} \geq (-\text{DNTwait}^n_{lpl'q}) \qquad (n,l,l') \in TP, p \in P_l, q \in Q^*_{l'} \tag{88}$$

The objective function, expression (85), includes all the components of the original model's objective function plus an extra summation over a new set of variables, (85e). The new variable $\texttt{ADNTwait}^n_{lpl'q}$ represents the absolute value of the difference between the transfer waiting time resulting while solving the UM model, $\texttt{NTwait}^n_{lpl'q}$, and the given waiting time, $\texttt{ONTwait}^n_{lpl'q}$, from the obtained infeasible solution via the Lagrangian-relaxation method. All of the constraints of this model are the same as the original model. However, we need to also calculate the value of `ONTwait` and `ADNTwait` variables. Constraints (86)-(88) are defined for this aim. Note that the `UM` model, is not node-based, since we want its solution to be a feasible solution to the original model. In other words, it includes constraints (6), (33), and all the capacity-related ones. Accordingly, the `UM` model aims to generate a truly feasible timetable that does not deviate significantly from the one suggested by the relaxation solution in terms of passenger's wait times. In other words, it strives to find an upper bound close to the best known lower bound.

6.2. Lagrangian Multipliers Update Mechanism

As mentioned above, the values of Lagrangian multipliers impact the `LR` model solution and in turn the feasible solution resulting from the `UM` model. We apply the subgradient algorithm to improve the Lagrangian multipliers to obtain the best possible lower- and upper-bound values and the closest possible feasible solution to the optimal solution of the original model. The subgradient algorithm is an iterative procedure that updates the Lagrangian multipliers in each iteration based on the obtained solutions up to that iteration. The solutions in our case are the upper- and lower-bound values and the timetables obtained from the subproblems.

More specifically, to update Lagrangian multipliers, we use expression (90) which has been used in previous studies (Zhang, Gao, Yang, Gao and Qi, 2020; Yin, Yang, Tang, Gao and Ran, 2017; Yang and Zhou, 2014; Caprara, Fischetti and Toth, 2002). The sg_{lp}^{nk} is an element from the subgradient vector calculated using the optimal solution of subproblems in iteration k, i.e., solving the subproblems with given values of the Lagrangian multipliers of iteration k. Subgradients are determined by expression (89) in which the $\texttt{arr}_{lp}^{nk}$ and $\texttt{dep}_{lp}^{n_{prev}k}$ are from different subproblems' optimal solutions since they do not belong to the same transfer node. In expression (90), the `UB` and `LB` denote the best upper- and lower-bounds up to iteration k. Note that, the `UB` and `LB` are respectively the best upper and lower bounds obtained across all iterations up to iteration k and do not necessarily come from the previous iteration, $k-1$.

$$sg_{lp}^{nk} = \text{arr}_{lp}^{nk} - \text{dep}_{lp}^{n_{prev}k} - ttd_{l}^{n_{prev}-n} \qquad l \in L_b^n, (n_{prev}, n) \in TN_l \setminus \{n_1(l)\}, p \in P_l \tag{89}$$

$$\lambda_{lp}^{n(k+1)} = \lambda_{lp}^{nk} + \frac{|UB - LB|}{||sg_{lp}^{nk}||^2} sg_{lp}^{nk} \qquad l \in L_b^n, p \in P_l \tag{90}$$

Algorithm 1 Updating Lagrangian multipliers

Initialize	Define the termination criteria: relative convergence tolerance of θ $k \leftarrow 0, \lambda_{lp}^{nk} \leftarrow 0,\ \ l \in L_b^n, p \in P_l$
Step 1:	Solve the subproblems, saving all SV^{nk}, arr_{lp}^{nk}, and $\text{dep}_{lp}^{n_{prev}k}$ values Calculate sg_{lp}^{nk} using (89) and determine the UB^k and LB^k
Step 2:	Compare UB^k and LB^k with the best UB and LB, and update the best bounds if needed
Step 3:	Determine the termination gap $G^k = (UB\text{-}LB)/UB$
Step 4:	If G^k less than θ terminate, otherwise go to the next step
Step 5:	$k \leftarrow k+1$ and update λ_{lp}^{nk} using (90)
Step 6:	Go to Step 1

The algorithm for updating the Lagrangian multipliers is displayed in Algorithm 1. We initialize the λ_{lp}^{nk} values as zeros and then solve the subproblems to start the iterations. The solutions of the subproblems are the optimal timetables for each node without any consideration of other nodes, even indirect effect, since in the first iteration the λ_{lp}^{nk} values are all zeros. Then, based on these solutions, the first feasible solution is achieved through the `UM` optimization model. The loop for updating the Lagrangian multipliers stops when a stopping criterion is met. This criterion can be the number of iterations, a given time budget, or when the difference, G^k, between the best obtained upper and lower bounds, i.e., `UB` and `LB`, is less than a target value, θ. If the gap is acceptable (or equals zero), meaning that the solution of the `UM` model is close enough to the optimal solution of the original model, we obtain the final solution of the model. Nevertheless, as we will observe in our experiments (Section 7.2), the solution of the `UM` model can be used as a warm start for a commercial solver, such as GUROBI, to solve our original MIP model to optimality. In fact, our solution method not only determines the optimality gap but also provides an effective warm start for commercial solvers to solve the model to optimality more efficiently.

7. Numerical Experiments

In this section, we examine our proposed model and solution method through various numerical experiments. We choose two three-node networks from the City of Toronto Transit Network (CTTN) for the first experiment. As we discuss later in Section 7.1, we compare different versions of our model with two other timetabling approaches: (1) a common conventional transfer time optimization model in the literature, model *CM* (Conventional Model), and (2) the regular practice of timetabling by agencies, model *NT* (No Transfer synchronization). We conduct this experiment to illustrate the necessity and benefits of considering the detailed process of a successful transfer, including dwell time determination and bus capacity consideration, particularly when transfer buffer times are involved. In the second experiment (Section 7.2), we examine different sub-networks with five, seven, and nine transfer nodes from the CTTN to showcase how our solution method performs for medium-size networks. Our solution method can solve different instances efficiently, while GUROBI has limited ability to provide bounds during the same time budget as our model. In these two experiments, we do not consider headway irregularity and bunching incidents, constraints (78)-(80). However, in Section 7.3, where we conduct a case study on a network with 12 transfer nodes from the CTTN as illustrated in Figure 2, it is important to assess the impact of headway irregularity and bunching penalty on the final solutions, i.e., generated bus timetables. The characteristics of sub-networks used in the following experiments, mainly the chosen transfer nodes and the associated headway combination of intersecting lines, the number of transfer directions, and total transferring demand during planning are shown in Table 5.

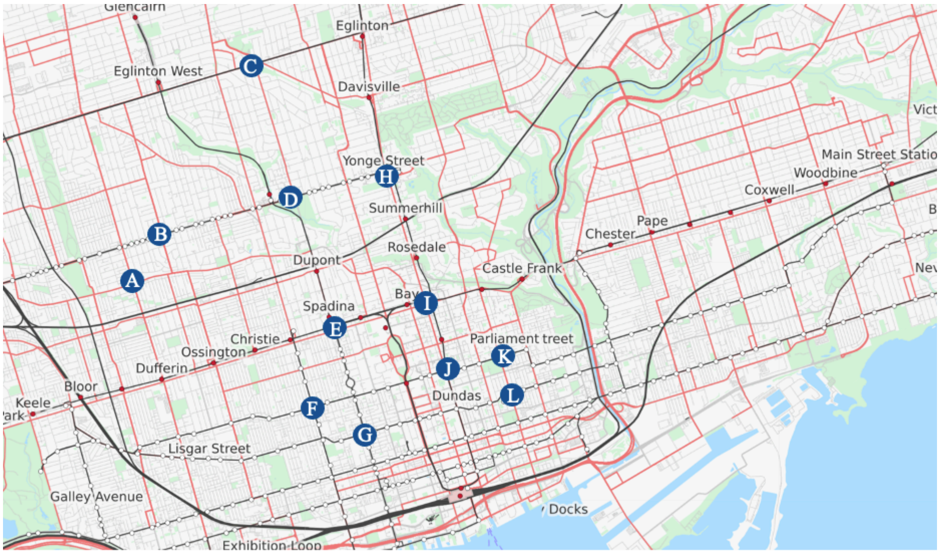

Figure 2: 12 Transfer Nodes in Downtown Toronto

Applying reasonable coefficients for different waiting time components in the objective function is essential because these coefficients heavily influence the credibility of solutions. Thus, we use the various waiting time weights derived in the literature of passengers' time perceptions (Gavriilidou and Cats, 2019; Cats, Rufi and Koutsopoulos, 2014; Cats, West and Eliasson, 2016): in-vehicle waiting time (c^{vt}= 1.5), transfer waiting time (c^{tw}= 2), penalty of first leftover passenger (c^{ft}= 7), penalty of second leftover passengers (c^{st}= 7). Note that these values are not explicitly calibrated for transit users of the CTTN. In fact, these passengers' time perception coefficient values might differ among cities and transit systems. These specific coefficients are not available for CTTN transit users to the best of our knowledge. Nevertheless, since these parameters are determined independently from a timetabling model, they do not affect our experiments and comparative analysis.

We produced the required data sets for these examples and the case study from a simulation platform that mimics real-world transit operations through an agent-based approach. This mesoscopic simulation platform, named Nexus, has been developed at the University of Toronto (Srikukenthiran and Shalaby, 2017). The input data sets to this simulation are General Transit Feed Specification (GTFS) data and travel behavior survey data. We use the simulation

Table 5
Characteristics of Chosen Transfer Nodes from the CTTN

Network Name	Transfer Nodes	Number of Routes	Min, Max, and Average Headway (min)	Number of Transfer Directions	Total Transfer Demand
3NI	K-L-J	10	(4.6) - (30.0) - (9.7)	20	349
3NII	E-G-L	14	(5.0) - (17.0) - (7.9)	24	438
5NI	K-F-J-G-L	18	(4.0) - (30.0) - (8.0)	44	842
5NII	H-D-J-K-B	10	(2.8) - (30.0) - (12.8)	15	229
5NIII	E-A-G-L-K	22	(4.6) - (17-8) - (12.0)	41	683
7NI	J-F-G-L-E-K-H	22	(2.8) - (30.0) - (9.2)	47	886
7NII	J-F-G-L-E-K-D	23	(2.8) - (30.0) - (8.2)	47	864
7NIII	E-A-L-K-D-H-J	21	(2.8) - (30.0) - (11.2)	32	478
9NI	J-F-A-G-L-K-E-H-D	29	(2.8) - (30.0) - (9.7)	56	971
9NII	J-F-G-L-E-K-H-B-I	22	(2.8) - (30.0) - (9.2)	47	886
9NIII	C-E-F-L-K-D-H-J-I	21	(2.8) - (30.0) - (11.2)	38	673
12N	K-E-I-G-F-B-D-H-C-J-L-A	32	(2.8) - (30.0) - (10.1)	59	1004

model of the City of Toronto, based on the data collected in 2016, for the morning peak hours 6-9 a.m. Also, for two other parameters of alighting time and boarding time per passenger, we use the values estimated for the CTTN, which are 1.12 and 1.96 seconds per person, respectively (Miller, Vaughan, Yusuf and Higuchi, 2018).

All the methods are implemented using Python and all the mathematical models are solved by GUROBI (Gurobi Optimization LLC, 2020), on a MacBook Pro with a 2.8 GHz (Intel Core i7) processor and 16 GB memory.

7.1. Experiment One: Three-node Networks

In this experiment, we compare the `CM` and `NS` models with three versions of our model: (1) model `NB` where no transfer buffer time, `TB`, is allowed; (2) model `WBTI` where `TB` is allowed with an upper limit on its value for each bus at each transfer node (e.g., we consider 10% of the line's headway); and (3) model `WBTII` where `TB` is allowed but the total `TB` value for each bus through all transfer nodes is less than a specific value. In both the `CM` and `NS` models, bus dwell times are given as inputs, and no transfer buffer times and capacity limitations are taken into account. Furthermore, in these two models, a successful transfer event is defined when the arrival time of transferring passengers at the connecting bus stop is less than the departure time of their connecting bus. On the other hand, in the model `NS`, the common timetabling practice followed by many agencies is applied, treating each line individually without considering transfer synchronization among lines. The timetables for model `NS` are derived from the Nexus simulation which are based on static GTFS data representing the timetables generated by agencies.

For this experiment, we choose two networks with three transfer nodes from the CTTN, namely 3NI and 3NII. To compare the solutions, we consider different performance metrics: total trip synchronization waiting times (transfer waiting time without demand consideration), total transferring passenger waiting time (transfer waiting time multiplied by the number of passengers who would experience that waiting time interval), average transfer waiting times (weighted by the number of passengers who would experience each waiting time interval), total dwell times, total transfer buffer times assigned to all buses, maximum transfer buffer time that occurs in our solution, total number of passengers experiencing transfer buffer times, total value of each transfer buffer time multiplied by the number of passengers who would experience that time, and average transfer buffer time experienced per passenger. As explained in Section 5, for feasibility purposes, a connecting bus from the next horizon can also be assigned as a transfer connection for a feeder bus. For the transferring passengers who are assigned to the next planning horizon, we assume two times of their connecting line's headway as the transfer waiting time. This is based on the assumption that lines are less frequent in off-peak periods. So if a metric is considering these passengers, it is indicated as `WNH`, With the Next Horizon.

7.1.1. *Preparing and evaluating the solutions of `CM` and `NS` models for the comparative analysis*

In order to compare the solution quality of the `NS` and `CM` models with our models, we need a transfer time calculation tool. This tool evaluates the generated timetables by the NS and CM models in a framework similar to our model with more details on the required dwell times and bus capacity. The input to this tool is bus arrival times at their first transfer node recommended by a model. For instance, the arrival times of buses at their first transfer node designed by the CM model are given to the calculation tool. Then, instead of using fixed values, the tool determines dwell times based on the number of boarding and alighting passengers and bus capacity, similar to our models. Note that a bus dwell time value determined through the calculation tool may differ from its corresponding fixed value. This difference shows the importance of determining accurate dwell times in models rather than using fixed values. This calculation is close to what would happen in the real world if the timetables from the CM model were applied. Furthermore, in the calculation tool, the definition of a successful transfer is similar to what we have considered in our models, i.e., Types 1 and 2. Type 3 of successful transfers is not considered since no transfer buffer is involved in the CM and NS models. To prepare the solutions of the CM and NS models for evaluation, we use our NB model with some minor changes as the calculation tool; since the NB model has the main transfer process and dwell time determination formulations while also considering the bus capacity. So if we fix the arrival times of buses at their first node based on given timetables of the CM model, then the NB model determines dwell times of buses, successful transfers, and transfer waiting times. The same steps are applied for the NS model in which the input timetables for the calculation tool are obtained from the Nexus simulation.

7.1.2. Results of experiment one

The distribution of transfer waiting times versus the number of passengers who would experience those times are shown in Figure 3 and Figure 4 for the 3NI and 3NII networks, respectively. In these figures, we eliminate the successful transfers that occur in the next planning horizon and are common in the solutions of all the models. In other words, if all models fail to assign a successful connection to a feeder during the planning horizon, those transfer waiting time records are not included in Figure 3 and Figure 4.

Nevertheless, we discuss and refer to the out-of-horizon waiting times in Table 6 and Table 7. Although our models attempt to assign as many passengers as possible with the least transfer waiting time during the given planning horizon, our scope does not consider issues regarding connections for the last buses within the horizon. For more information on transfer synchronization for the last buses in a planning horizon, we refer the readers to some other studies (Long, Meng, Miao, Hong and Corman, 2020; Guo, Wu, Sun, Yang, Jin and Wang, 2020).

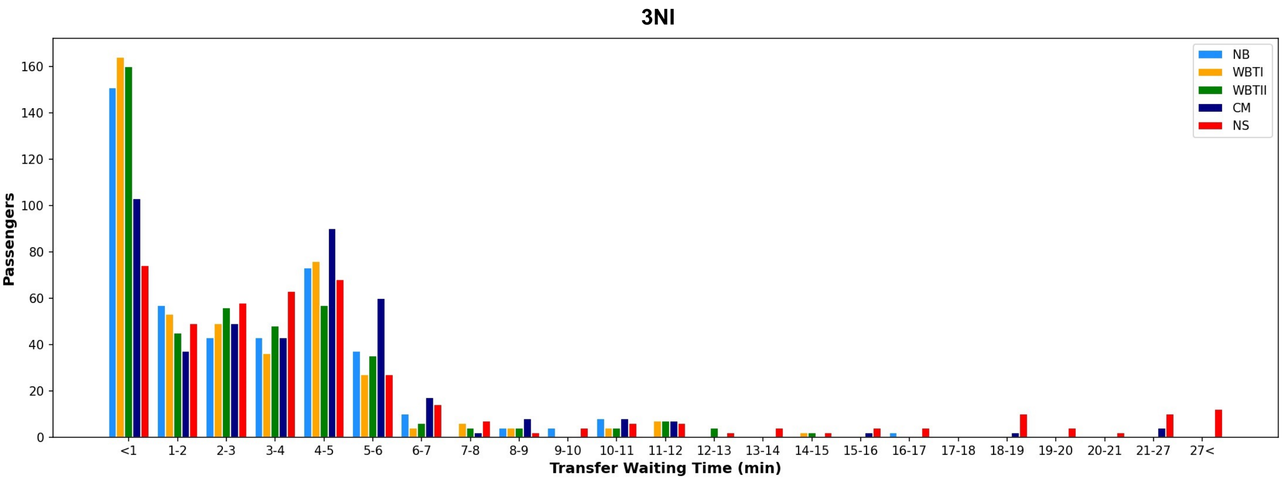

Figure 3: Model Comparison: 3NI

As shown in both Figure 3 and Figure 4, the distributions of transfer waiting times are skewed to the left, meaning that the majority of passengers would have short transfer waiting times in all models' solutions. However, the number of very short transfer waiting time records, less than one minute, is considerably higher in the solution of the NB,

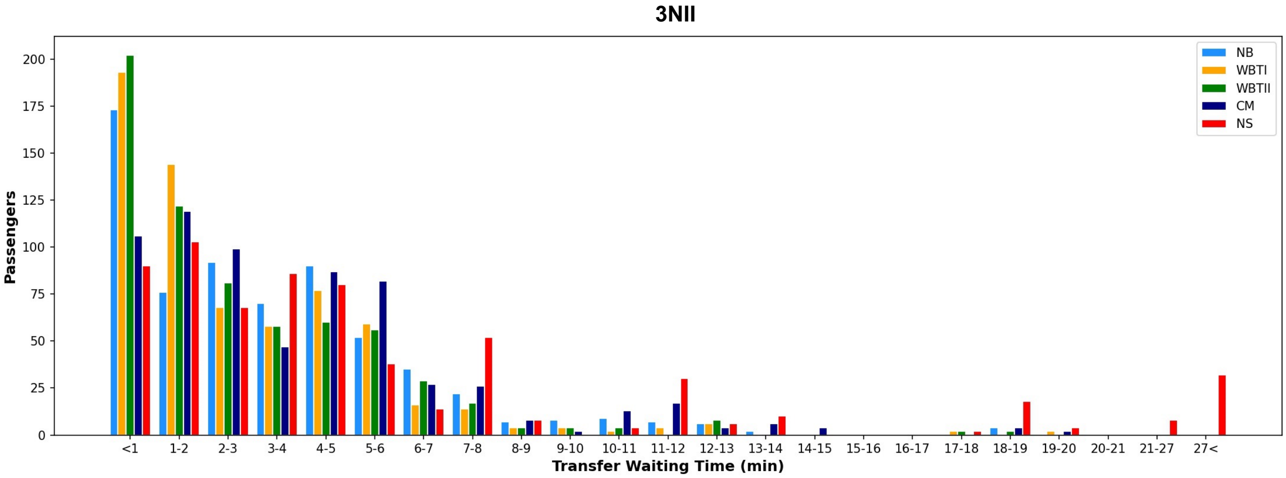

Figure 4: Models Comparison: 3NII

WBTI, and WBTII models compared to the CM and NS models. In other words, the number of passengers who would experience short transfer waiting times in our proposed timetables is higher compared to the timetables obtained form the models CM and NS. This superior performance is likely due to two primary reasons. First, our detailed model captures and represents a more realistic synchronization procedure, including dwell time determination and bus capacity consideration. Even though the model NB has no transfer buffer time, the number of cases with zero transfer waiting times is higher compared to the models CM and NS. Second, the use of transfer buffer times enables the WBTI and WBTII models to achieve more zero transfer waiting time records and almost no long ones, since the addition of very short TB can transform long transfer waiting times to zero waiting times. Nevertheless, since the model also has to consider restrictions while introducing the transfer buffer time, it is not always possible to avoid long transfer waiting times. For instance, if a bus with many on-board passengers and the next feeder bus is expected to have a small number of transferring passengers, or if there is limited spare capacity for passengers to board, the model may decide not to hold the bus, i.e., assigning no transfer buffer time. In this case, long transfer waiting times can occur in our timetables, as shown in Figure 3 and Figure 4. As mentioned before, adding transfer buffer times to reduce transfer waiting times has some trade-offs involved. In addition to the extra in-vehicle times for already on-board passengers, including through and Types 1 and 2 successfully transferred passengers, and a potential increase in bus cycle times, it is also possible that some short transfer waiting times convert to longer ones as a result of adding transfer buffer times. The increase in some transfer waiting time records occurs when our WBTI and WBTII models reduce some waiting time records, either transfer or in-vehicle waiting times, and increases some others to minimize the total objective value. However, to control this behavior to some extent in our problem, we consider demand in the objective function to guide models to focus on both the value of each type of transfer/in-vehicle waiting times and their associated demand simultaneously. We illustrate how the benefits of allowing transfer buffer times exceed its possible drawbacks in Figure 5 and Figure 6. In these figures, we show the determined transfer waiting times, for each transfer pair of WBTI model on the x-axis and the CM model on the y-axis. The bubble size indicates the number of passengers experiencing those times. One can see how our WBTI and WBTII models not only reduce the total transferring passenger waiting times, which will be discussed later in Table 6 and Table 7 but also how they decrease the very long transfer waiting time records in particular. For instance, in Figure 5 some long waiting times resulting from the model CM are reduced to zero in the model WBTI. The same pattern is seen in Figure 6. Nevertheless, inevitably, as discussed earlier, some short transfer waiting times in the model CM have increased to longer ones in the model WBTI.

Note that although the models' performances are similar in both 3NI and 3NII, the combination of headways and demand are different in these two sets, leading to different opportunities and possibilities for our models in terms of reducing total and long transfer waiting times. The average, maximum and minimum headways in 3NI are 9.7, 30.0,

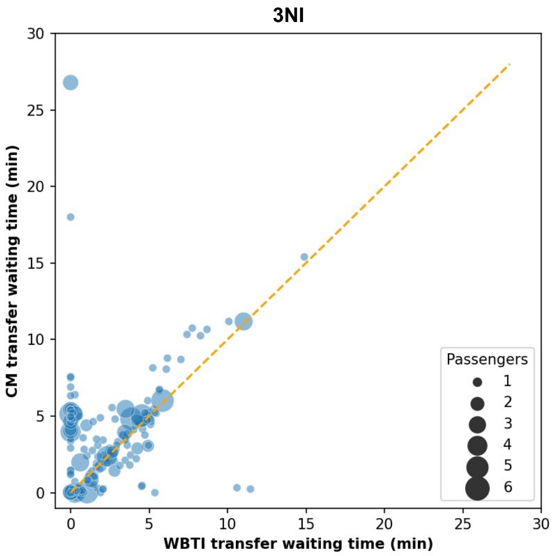

Figure 5: `WBTI` and `CM` comparison: 3NI

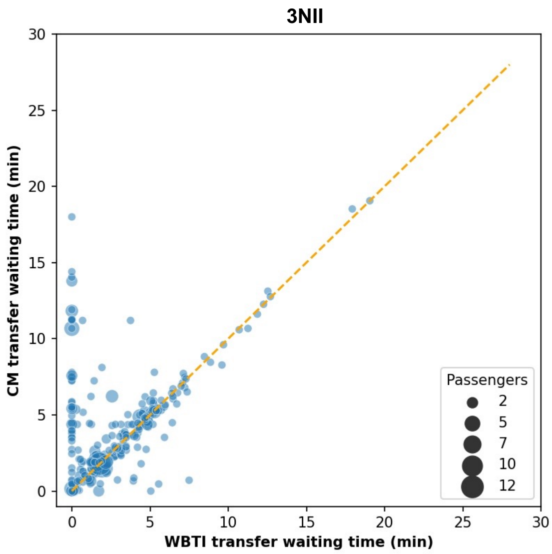

Figure 6: `WBTI` and `CM` comparison: 3NII

4.6 minutes, while for 3NII they are 8.1, 17.0, and 5.0 minutes, respectively. In 3NI, where the difference between the maximum and minimum headway is large, `WBTI` and `CM` have more similar records, i.e., transfer waiting times on the diagonal line, compared to 3NII. The reason is that in 3NI with maximum headway of 30 minutes, the model `WBTI` mostly focuses on reducing waiting times corresponding to transfers to low-frequency lines. Otherwise, those waiting times would end up being very long. On the other hand, for 3NII, the maximum waiting time can be 17 minutes, which equals the maximum headway. So, the model focuses on other waiting times to reduce, e.g., transfers to medium-frequency lines with 10-minute headway, more than transfers to low-frequency lines and long transfer waiting times. Therefore, as shown in Figure 3 and Figure 4 as well as Figure 5 and Figure 6, the frequency of short transfer waiting times, less than two minutes, is considerably higher for 3NII. That being said, as discussed above, the associated demand of each pair of transfers also affects how the model optimizes the total value of the objective function.

The comparison of the various models with respect to key metrics are presented in Table 6 and Table 7 for 3NI and 3NII, respectively. The percentages in parentheses show the improvement of each model compared to the `NS` model for that specific metric. As expected, the `CM` model's improvement is the least. The considerable improvement in the performance of the model `NB` shows the distinct advantage of our successful transfer formulation, since even without the aid of transfer buffer the `NB` model could reduce the total transferring passengers waiting time, not including the next horizon connections, by 50% and 41% in 3NI and 3NII, respectively.

The improvement observed in the solutions of the `WBTI` and `WBTII` models in terms of total transfer waiting times and total trip synchronization waiting times are almost the same in each 3NI and 3NII sub-networks individually. However, when these two models are applied for different networks, i.e., 3NI and 3NII, different patterns can be observed in the results. For instance, by using the model `WBTI` for 3NI, the total transfer buffer, maximum transfer buffer, and average transfer buffer penalty that would be experienced by each passenger are 5.26, 1.20, and 0.42 minutes while for 3NII these metrics are 20.39, 1.70, and 0.58 minutes, respectively. As mentioned above, the reason for this inconsistent execution of transfer buffer times in different networks by the same model depends on several factors, such as headway combinations and demand distributions. Note that, as explained earlier, we did not consider the exact value of `TB` as a penalty in the objective function in order to prevent non-linearity. Instead, based on the solutions, we determine the actual penalty of each assigned `TB` experienced by the associated passengers. Then based on the total number of transfer buffer experienced by passengers, the average transfer buffer penalties are calculated.

The differences in values of the metrics coded with `WNH` and the associated values without `WHN` explicitly reveal how

Table 6
Metric comparison for different models, 3NI

Metric	NB	WBTI	WBTII	CM	NS
Total trips' synchronization waiting times (min)	489	470	469	627	1019
	(↓52%)	(↓54%)	(↓54%)	(↓38%)	—
WNH total trips' synchronization waiting times (min)	1193	1109	1108	1319	1590
	(↓25%)	(↓30%)	(↓30%)	(↓17%)	—
Total transfer waiting time * demand (person·min)	672	640	698	948	1350
	(↓50%)	(↓53%)	(↓48%)	(↓30%)	—
WNH total transfer waiting time * demand (person·min)	1674	1577	1635	1939	1959
	(↓15%)	(↓20%)	(↓17%)	(↓1%)	—
Average transfer waiting time for a passenger (min)	1.93	1.83	2.00	2.70	3.87
WNH average transfer waiting time for a passenger (min)	4.79	4.52	4.69	5.55	5.61
Total dwell time (min)	38.07	43.00	44.17	38.23	38.93
Total transfer buffer (min)	0	5.26	6.23	0	0
Maximum transfer buffer (min)	0	1.2	1.2	0	0
Transfer buffer experienced passengers	0	169	88	0	0
Total transfer buffer * experienced passengers (person·min)	0	71.1	45.9	0	0
Average transfer buffer penalty (min)	0	0.42	0.52	0	0

some transferring passengers are assigned to buses operating in the next planning horizon. The number of passengers assigned to the next horizon was almost the same for all the models in each set. However, in our models' solutions, these passengers have shorter transfer waiting times. The reason is that the `WBTI` and `WBTI` models assign passengers to the next horizon with connections mostly to a high frequency to minimize the total objective value. Moreover, they reduce the transfer waiting times of connections to low-frequency with the aid of transfer buffer times to avoid large transfer waiting times occurring in the next planning horizon, as also seen in Figure 5 and Figure 6.

The number of leftover passengers is also the same for all models in both sets. The value of leftover passengers shows how a system with current resources, such as the number of trips and bus sizes, is operating in terms of spare bus capacities at the vehicle level. As discussed above, line frequency setting is usually based on hourly demand profiles at stops rather than the number of boarding and alighting passengers for each trip. Therefore, bus capacity constraints provide agencies with insights regarding whether the chosen frequencies and bus sizes in the previous planning stages are suitable for serving passengers. On the other hand, the main reasons for taking into account the bus capacity constraints in our formulation are to facilitate more realistic bus dwell time determinations and, accordingly, more accurate bus trajectories and transfer waiting time calculations to the extent possible.

Note that the restrictions for the maximum allowable values of buffer times are the same in both sets and can be adjusted based on preferences of agencies. Although the total buffer times added to bus cycle times are larger in 3NII, they are still less than the 10% percent of a line's headway that we enforced in these experiments. Moreover, the total buffer time of each line should be considered individually to investigate the possible extra cost for agencies. For instance, if the introduced transfer buffer time to a trip can be included in its terminal times, then there is no substantial additional cost for agencies. Otherwise, constraints regarding the `TB` limits can be defined in a way to combine the `WBTI` and `WBTII` models in order to take into account preferences by agencies.

7.2. Experiment Two: Networks of Five, Seven, and Nine Transfer Nodes

To investigate the performance of our solution method, we choose nine networks, with five, seven, and nine transfer nodes from the CTTN. In this experiment, we only compare the solution quality of the `WBTI` and `CM` models. To solve our model using the Lagrangian relaxation approach, we implement Algorithm 1 where as the stopping criterion, we choose θ equals 15%, i.e., we stop when the difference between the best obtained lower and upper bounds is less than 15% of the upper bound. Next, we use the feasible solution that provides the best obtained upper bound as a warm start

Table 7
Metric comparison for different models, 3NII

Metrics	NB	WBTI	WBTII	CM	NS
Total trips' synchronization waiting times (min)	961	799	796	1036	1817
	(↓47%)	(↓56%)	(↓56%)	(↓43%)	—
WNH total trips' synchronization waiting times (min)	1666	1464	1461	1834	2533
	(↓34%)	(↓42%)	(↓42%)	(↓28%)	—
Total transfer waiting time * demand (person·min)	1243	992	978	1377	2100
	(↓41%)	(↓53%)	(↓52%)	(↓34%)	—
WNH total transfer waiting time * demand (person·min)	2073	1782	1767	2321	2837
	(↓27%)	(↓37%)	(↓38%)	(↓18%)	—
Average transfer waiting time for a passenger (min)	2.84	2.26	2.23	3.14	4.80
WNH average transfer waiting time for a passenger (min)	4.73	4.07	4.03	5.30	6.48
Total dwell time (min)	41.82	61.88	59.35	41.87	42.17
Total transfer buffer (min)	0	20.39	17.72	0	0
Maximum transfer buffer (min)	0	1.7	1.7	0	0
Transfer buffer experienced passengers	0	188	193	0	0
Total transfer buffer * experienced passengers (person·min)	0	108.9	119.6	0	0
Average transfer buffer penalty (min)	0	0.58	0.62	0	0

to GUROBI and let it run for one hour. Note that we can either adjust the stopping criterion to a smaller acceptable gap, $\theta = 0$, or let the model, with the warm start, run for longer than one hour to achieve optimality for all networks. However, in this experiment, we obtained high-quality solutions for our comparison purposes without spending extra time to solve the models to optimality. We compare the performance of our solution method with that of solving our model directly with GUROBI in terms of computational time as well as lower- and upper-bound improvements. To do so, we run the models using GUROBI with and without a warm start. Finally, we assess the quality of our solutions with those of the `CM` model.

The first lower bound with all λ^n_{lp}'s equal zero, the first upper bound, and the best obtained lower and upper bounds across all iterations through implementing Algorithm 1, are shown in Table 8 in percentage values. For each network, the percentages show the difference of each bound from the best solution objective value, obtained by running the model using GUROBI with a warm start for one hour. For instance, for set 5NI, the best lower bound is 3,683,363, and the final accepted objective function value is 4,250,379, so the best lower bound is 13.34% of 4,250,379, (15.00%-13.34% = 1.66%) smaller than the final accepted objective function value. Clearly, the percentages for upper bounds show how much those values are larger than the final accepted objective function value. We use the same approach for the other metrics. As reported in Table 8, the best upper bounds obtained via Algorithm 1 as our solution method are very close to the true optimal value for all the networks. The farthest value is for network 7NIII with a 5.21% difference from the final solution. This comparison confirms that we have achieved highly promising solutions from the algorithm with the best obtained upper bound without the need to run the model using GUROBI with a warm start for another one hour.

In this experiment, we compare our solution method's computational time and performance with the GUROBI solver to demonstrate the efficiency of our Lagrangian relaxation-based approach. To do so, we run the `WBTI` model using GUROBI with no warm start (indicated by `NWS`) for the total duration of implementing Algorithm 1 to update the Lagrangian multipliers and reach the stopping criterion plus the extra one hour for each network of transfer nodes. One can see that the performance of GUROBI is rather poor compared to our solution method with the same time budgets. For example, for 5NIII and 9NI networks, GUROBI could not even provide any upper or lower bound in a reasonable time while our solution method solved the 5NIII network to optimality and the 9NI network to 5% gap. Thus, the results demonstrate that the GUROBI solver applied to our problem performs inadequately, showing the need of an efficient solution method for our model. The results also indicate the strong potential of our solution method owing to

Table 8
Solution method performance of WBTI model for different networks

Metric	5NI	5NII	5NIII	7NI	7NII	7NIII	9NI	9NII	9NIII
First lower-bound	13.34%	19.40%	21.17%	16.09%	12.02%	12.45%	13.43%	13.86%	7.09%
First upper-bound	12.29%	219.82%	271.18%	53.08%	18.07%	101.37%	54.10%	51.26%	68.39%
Best lower-bound	13.34%	12.25%	14.67%	12.76%	12.02%	9.32%	10.96%	13.35%	7.10%
Best upper-bound	0.00%	0.07%	0.08%	0.00%	0.00%	5.21%	4.59%	0.84%	5.48%
Number of iteration	2	12	18	5	2	26	15	6	2
Iterations time (min)	10.76	1.29	1.65	17.29	10.23	17.78	43.92	18.18	4.81
GUROBI starting gap	7.84%	2.41%	1.53%	9.68%	1.55%	6.19%	11.3%	8.35%	5.31%
GUROBI gap after 1h	4.86%	0.00%	0.00%	7.37%	0.00%	0.00%	4.97%	0.82%	0.03%
GUROBI gap NWS	15.00%	0.00%	NA	13.10%	9.78%	0.02%	NA	14.70%	0.03%

Table 9
Metrics comparison for different networks (average values), models WBTI and CM

Metrics	5Nodes	7Nodes	9Nodes
Total vehicle waiting time (min)	↓6.47%	↓11.88%	↓11.75%
WNH total vehicle waiting time (min)	↓10.70%	↓12.25%	↓12.97%
Total transfer waiting time * demand (person·min)	↓9.22%	↓13.45%	↓14.81%
WNH total transfer waiting time * demand (person·min)	↓12.08%	↓12.54%	↓14.27%
Average transfer waiting time (min)	↓12.08%	↓12.54%	↓14.27%
WNH average transfer waiting time (min)	↓12.08%	↓12.54%	↓14.27%
Total dwell time (min)	↑20.56%	↑16.26%	↑27.33%
Total transfer buffer (min)	18.34	13.61	26.14
Maximum transfer buffer (min)	2.24	2.35	2.94
Average transfer buffer penalty (min)	0.81	0.82	0.94

its promising performance. Note that in some 0.00% gap cases, for instance, 5NII, the model achieved optimality in less than one hour.

Additionally, our experiments for networks with different numbers of transfer nodes confirmed that the number of transfer nodes is not the only factor impacting computational burden. In fact, how transfer nodes are connected to one another, the number of transfer nodes in a line's sequence, headway combinations, and demand distribution also considerably affect =the required time for solving the models. For instance, as presented in Table 8, for the 7NI network, the computational time increased compared to 5-node networks. On the other hand, for the 9NIII network, the number of iterations and total computational time are noticeably less than those of all 7-node networks.

In order to further examine the quality of the final solutions, we examine the above-mentioned performance metrics in experiment one for the WBTI and CM models. The percentages in Table 9 indicate the improvement in the WBTI model compared to the CM model over the associated value in the CM model. Note that we present the average values for each set of networks in Table 9. Although we did not solve our models to optimality, the results show that even with around a 5% gap, our models performed considerably better than the model CM. Our model reduced the total passenger transfer waiting times by almost 14% for the 9-node networks and substantially decreased the average transfer waiting times per passenger. While, on average, we only added 0.81, 0.82, and 0.89 minute transfer buffer penalties per passenger in the 5, 7, and 9-node networks, respectively. Moreover, as illustrated in Table 9 the more the number of transfer nodes is being considered, the further improvement can be seen in the performance metrics. For instance, the average reduction in total transfer waiting times is almost 5% more for 9-node networks than 5-node networks. The main reason is that in subnetworks with a higher number of transfer nodes, while the negative effect of less-synchronized routes might

Table 10
Solution method performance of `WBTI` model for 12-nodes networks

Metric	`12NodesW`	`12Nodes`
First lower-bound	3,757,736	3,802,411
First upper-bound	6,997,169	6,969,166
Best lower-bound	3,899,792	3,935,919
Best upper-bound	4,541,737	4,584,488
Number of iteration	10	18
Total computation time for iterations (min)	37.43	62.51
Starting gap GUROBI with warm start	10.10%	10.30%
Best GUROBI objevtive value with warm start after 1 hour	4,541,462	4,553,943
GUROBI with warm start gap after 1 hour	4.84%	4.44%

be more severe, a larger number of transfers would occur during the planning horizon, particularly to low-frequency lines, leading to more opportunities to use transfer buffer times. Therefore, by applying our formulation and with the aid of transfer buffer times, long transfer waiting times in larger networks can be reduced to short waiting times more effectively.

7.3. Case Study

The majority of previous transfer synchronization studies have not examined headway regularity deviations at the scheduling stage. The main reason is that by considering fixed dwell time, scheduling trips to depart from terminals based on their fixed headways would result in regular headway-based arrival/departure at the following nodes. However, by determining the required dwell time and assigning varying transfer buffer times in our case, the model provides an opportunity to evaluate and, if needed, enforce headway regularity at all the stops. This experiment mainly examines how considering headway regularity deviations and the associated penalty inflicted on local passengers would affect obtained solutions. We choose a 12-node network from the CTTN, with 32 routes, 382 trips, and 1004 transferring passengers during 6-9 am, as shown in Figure 2. The shortest and longest route headways are 2.8 and 30.0 minutes, respectively. To examine the effects of headway regularity violations on our model's solution and its computational time, we solve the `WBTI` model twice: with constraints (78)-(79), the model `12NodesW`, and without the mentioned constraints, the model `12Nodes`. The coefficient in the objective function of the model `12NodesW` is 2 for headway deviations, set based on the study by Wardman (2004). Similarly, we use the transfer waiting time calculation tool, as described in Section 7, which enables comparing the solution quality of the `CM` model with and without constraints (78)-(79) with the `12NodesW` and `12Nodes` models. In what follows, we first compare the solution method performance for the `12NodesW` and `12Nodes` models. Subsequently, we investigate the solutions of the two models in terms of the previously-defined performance metrics, e.g., total transfer waiting time, headway regularity violations, and bunching incidents.

Table 10 presents the implementation output of our Lagrangian relaxation-based solution method. The model considering headway regularity violations and bunching incidents, `12NodesW`, is faster in providing the best lower and upper bounds. As seen, a gap less than 15% between the best bounds through Algorithm 1 was obtained almost two times faster for the model `12NodesW`. However, after running the models using GUROBI with a warm start for one hour, both the `12NodesW` and `12Nodes` models achieved nearly the same gap. Furthermore, similar to experiment two, the obtained upper bounds of the models through Algorithm 1 are close to the final solution that results from solving the model using GUROBI with a warm start. Nevertheless, this pattern may not be the same if the models are applied for another set of transfer nodes. In fact, as discussed above, there are several factors affecting the computational time and the solution method performance. However, in both of the models, the solution method was able to find a solution that is quite close to the optimal in an acceptable time where GUROBI alone cannot provide any bounds within the same time budget.

We explore the solution quality of the `12NodesW` and `12Node` models with regards to the above-mentioned metrics by comparing the solution of each model with its corresponding solution derived from the `CM` model. For instance, the

Table 11
Metrics Comparison for 12-Node Networks, `WBTI` model

Metric	`12NodesW`	`12Nodes`
Total vehicle synchronization waiting time (min)	2611.13(↓10.96%)	2676.35.42(↓8.73%)
`WNH` total vehicle synchronization waiting time (min)	4697.46(↓11.95%)	4753.95(↓10.89%)
Total transfer waiting time * demand (person.min)	3242.20(↓9.84%)	3246.38(↓9.72%)
`WNH` total transfer waiting time * demand (person.min)	5508.47(↓10.85%)	5518.91(↓10.68%)
Average transfer waiting for a passenger (min)	3.229(↓9.84%)	3.233(↓9.72%)
`WNH` average transfer waiting for a passenger (min)	5.487(↓10.85%)	5.497(↓10.68%)
Total dwell time (min)	143.28(↑13.23%)	142.53(↑12.64%)
Total transfer buffer (min)	18.00	16.98
Maximum transfer buffer (min)	2.55	2.55
Transfer buffer experienced passengers	706	696
Total transfer buffer * experienced passengers (person.min)	392.50	375.47
Average transfer buffer penalty (min)	0.56	0.54

percentage values, shown in Table 11, for model `12NodesW` show the improvement of a metric by model `12NodesW` compared with the solution of model `CM` being evaluated through the transfer waiting time tool calculation that also includes the headway regularity violation and bunching incident constraints. In terms of the solutions' quality, as shown in Table 11, both models are performing better compared to the conventional model, `CM`. On average, most of the metrics are improved by 10%.

Additionally, we compare the solutions of the `12NodesW` and `12Node` models with each other, i.e., not considering the improvement percentages with respect to the `CM` model. Most of the metric values, such as total and average transfer waiting times, are similar for both the `12NodesW` and `12Nodes` models. Although the total transfer buffer times and the number of passengers who would experience the buffer penalty are slightly higher in the model `12NodesW`, the average transfer penalty per passenger is not quite different from its value in the model `12Nodes`.

Nevertheless, as experiment one illustrates, the total and average values are not informative enough, and disaggregated analysis is also required for informing the decision-making process. In this regard, we explore how considering headway irregularity penalty affects the transfer waiting time values. In Figure 7 and Figure 8, we cluster the waiting times of each transfer pair into three groups based on its connecting line's headway: (1) headways less than 10 minutes are high-frequency and coded with blue; (2) headways between 10 and 20 minutes are medium-frequency and coded with orange; (3) headways more than 20 minutes are low-frequency and coded with green. Note that previous studies considered various threshold values to distinguish transit lines into high- and low-frequency. The majority of studies (Ansari Esfeh, Wirasinghe, Saidi and Kattan, 2021; Mazloumi, Mesbah, Ceder, Moridpour and Currie, 2012; Delgado, Munoz and Giesen, 2012) recommended 10-minute headway as the limit to separate high- and low-frequency lines. However, in our study, we group the headways into three types, high-, medium-, and low-frequency, to provide a more detailed discussion. As demonstrated in Figure 7 and Figure 8, the performances of the `12NodesW` and `12Nodes` models are similar for transferring to high- and medium-frequency lines, blue and orange records, respectively. In particular, almost all the blue records, transfers to high-frequency lines, are identical in both models mainly because, as expected, transfers to high-frequency lines are easier to be synchronized without the assistance of transfer buffer times. In fact, in some cases, as obvious from the performance of the `CM` model, transfers to high-frequency lines are already well-synchronized without applying a transfer synchronization model. However, some orange records, transfers to medium-frequency lines, are increased from less than 2 minutes in the `CM` model to more than 10 and 15 minutes in the `12Nodes` and `12NodesW` models, respectively. Nevertheless, the `12NodesW` performs considerably better compared to the `12Nodes` model in reducing very long waiting time records of transferring to low-frequency lines, green records. For instance, while the model `12NodesW` reduced transfer waiting times of more than 25 minutes to zero, the model `12Nodes` reduced those records to around 20 minutes and increased the transfer waiting time from less than 4 minutes to more than 23 minutes.

Additionally, by looking at Figure 7 and Figure 8 separately, one can see that there are some blue and orange records

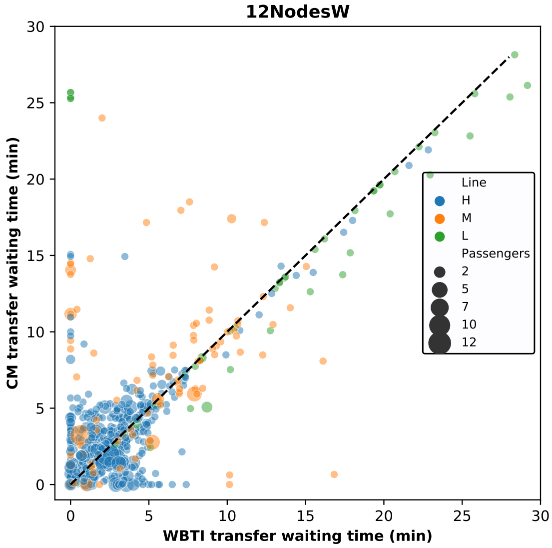

Figure 7: Models comparison: 12NodesW

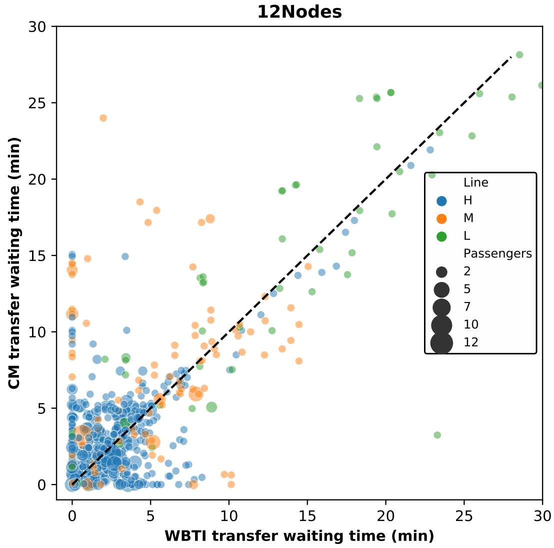

Figure 8: Models comparison: 12Nodes

with high values in the `CM` model which are reduced to zero in the `12NodesW` and `12Nodes` models. In contrast, some other short transfer waiting times in the `CM` model have been increased from zero to around 9.0 minutes in `12Nodes` and 7.5 minutes in `12NodesW`. The reason is that the focus of an optimization model heavily depends on the summation of its objective function components. Thus, our models, `12NodesW` and `12Nodes`, minimize the total value of the objective function in which some transfer waiting times might be sacrificed, i.e., increased, to reach the possible minimum of total transfer waiting times (and other types of waiting times in our case, e.g., extra in-vehicle time) for the whole system. As discussed in experiments one and two, for other long transfer waiting time values that the models cannot decrease them, mostly the ones on the diagonal line, there might be some other factors that prevent the model such as a high number of in-vehicle passengers leading to a negligible worth of transfer buffer times.

Comparing the metric values in Table 11 and transfer waiting time records in Figure 7 and Figure 8 we observe that considering headway regularity and bunching incidents may not considerably impact the solutions. In fact, considering constraints (78)-(79) and incorporating the violation components in the objective function of the `12NodesW` model result in slightly better solutions, i.e., smaller total and average transfer waiting time. The solution quality of the model `12NodesW` seems more favorable since enforcing the model to reduce headway regularity violations and bunching incidents supports the goal of minimizing transfer waiting times as well. Due to dwell time determination, the only way to improve the headway regularity is to assign transfer buffer time, i.e., hold a bus if needed. As explained in Section 5, we have a detailed model taking into account the individual components of bus dwell time as well as specifically formulating a successful transfer with the help of pre-planned holding times, Type 3. The extra transfer buffer time assigned in the `12NodesW` model improves not only the headway regularity but also the transfer synchronization leading to shorter transfer waiting times. We note that if the violations' weights in the objective function were larger than those chosen in this experiment, the `12NodesW` model might have a different solution. Needless to say, there is no single factor affecting the solution. So, if needed, agencies might need to run the model with different weights and different transfer buffer time limits to find a solution that satisfies their service standards and restrictions. Furthermore, note that these analysis and comparison outputs may not be the same for another set of transfer nodes.

Moreover, we investigate the values of headway regularity violations in the solutions of both the `12NodesW` and `12Nodes` models. The value of a headway regularity deviations equals the discrepancy in the interval times between the arrival times of two consecutive buses at a transfer node, i.e., headway based on the recommended timetables by a model, and the range of allowable minimum and maximum headways. However, this value alone is not necessarily

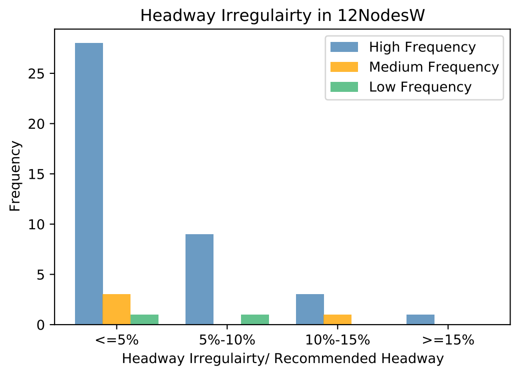

Figure 9: Models comparison: `12NodesW`

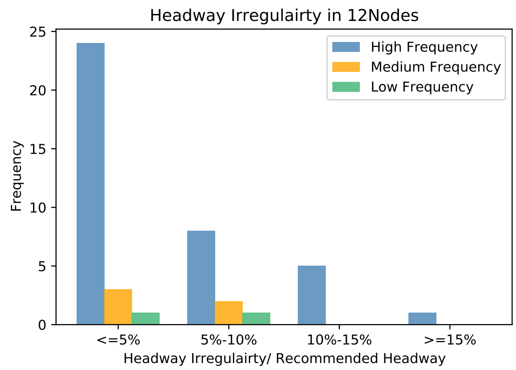

Figure 10: Models comparison: `12Nodes`

informative, since a 2-minute headway irregularity of low-frequency line with a 30-minute headway (and range of 27-33 minutes with allowable 10% headway variation) is less critical compared to a 2-minute violation for a high-frequency line with a 5-minute headway (allowable range of 4.5-5.5 minutes). Therefore, we divide the deviation values over their preferred given headway values, e.g., 2 minute /30 minutes. As illustrated in Figure 9 and Figure 10, for high-frequency lines, the number of headway regularity violations records for the (0-5%) category is more in Figure 9 while in Figure 10, the number of violations with (10%-15%) is larger. There are two main reasons why most violations have occurred for the high-frequency lines. First, the acceptable range of headways is smaller for high-frequency lines. For example, for a 10-minute headway, the allowable range is 9 to 11 minutes, which is more likely to be violated than a high-frequency line with a 30-minute headway and an allowable range of 27-33 minutes. Second, the number of high-frequency lines is more than medium- and low-frequency lines in our case study; over half of the 32 lines are high-frequency in our 12-node network.

Furthermore, as expected, we obtain fewer headway regularity violations from the `12NodesW` model compared to the `12Nodes` model for both medium- and high-frequency lines. Nevertheless, besides this difference, the behaviors of the two models with respect to headway irregularity are similar. Although assigning transfer buffer times contributes to two goals of reducing headway irregularities and transfer waiting times simultaneously, it also takes into account the induced penalty of extra in-vehicle times inflicted on the through passengers. In other words, the model might not assign buffer times for reducing headway irregularities and transfer waiting times if those buffer times would considerably increase the in-vehicle time penalties for on-board passengers. Note that in previous studies, headway regularity was not considered for all the transfer nodes since the main decision variables were the departure times of buses from terminals. Thus, no further analysis was examined to explore the arrival and departure times of trips happening at their following transfer nodes. However, we can consider headway regularity at all transfer nodes while quantifying the trade-off between headway regularity deviations and other critical performance metrics in timetabling and transfer synchronization, e.g., in-vehicle times and transfer waiting times, due to the detailed formulation of bus dwell time and bus capacity limit.

8. Summary and Conclusion

In this paper, we introduce a novel mixed-integer programming formulation for the transfer synchronization problem to mitigate the disutility associated with transferring by public transit users. We present an optimization framework for bus timetabling that is focused on minimizing transferring passenger waiting times by incorporating key details of the transfer process, particularly spare bus capacity and dwell time requirements. We track a connecting bus spare capacity at each transfer node, ensuring that a successful transfer actually can occur. Otherwise, even if feeder and connecting buses are well-synchronized, passengers might not be able to have a successful transfer and as a result wait for the next bus if spare capacity in the connecting bus was not available. Tracking bus capacities enables us to determine the required bus dwell times based on the number of alighting and successful boarding passengers. In lieu of fixed dwell time values in the timetabling formulation, determining the required dwell time for each bus, specifically in transfer synchronization models, enhances the reliability of timetables.

To examine spare bus capacity and determine suitable dwell times, we break down the dwell time at a connecting bus stop into three intervals based on the transferring passengers' arrival times at the bus stop relative to their connecting bus arrival time. For each interval, the amount of passenger demand, non-transferring and transferring passengers, is compared to the available capacity of the bus, and accordingly, the number of passengers that can successfully board and their required boarding time are determined. To explicitly include transferring passengers waiting time in the dwell time formulation, we propose setting three distinguished types of successful transfers. The first type includes passengers arriving before the arrival of their connecting bus - they clearly need to wait for their connecting bus. Second, passengers arriving on time, i.e., at the same arrival time of the connecting bus or shortly after - these passengers would not experience any transfer waiting time. For the third type, we introduce a new concept, namely transfer buffer times, in which pre-planned holding time is added to a connecting bus dwell time to improve transfer synchronizations while considering the extra in-vehicle time penalty of already on-board passengers, including through, local, and successfully transferred passengers Types 1 and 2. The third type is primarily beneficial for transferring to low-frequency lines, whereby holding a bus for a short duration transfer waiting times can be reduced to zero.

To solve our complex formulation, we design a Lagrangian relaxation-based solution method to efficiently solve the model for large-scale networks. Although this method provides a valid lower bound on the optimal objective value, the generated solution is not necessarily feasible. Thus, we propose a heuristic optimization model to transform the infeasible solution to a feasible one. The benefit of our solution method, aside from its efficiency and effectiveness compared to traditional heuristic approaches is the provided optimality gap. If the gap is small enough, the corresponding solution of the best obtained upper bound can be used as the final solution reducing the computational burden of providing an actual optimal solution.

We examine the performance of our model and solution method through three experiments on sub-networks from the City of Toronto Transit Network. In experiment one, we test three different versions of our model for two sets of three-node networks: (1) no transfer buffer time is considered, model `NB`; (2) transfer buffer time is added to each bus at each node with an upper limit, model `WBTI`; and (3) transfer buffer time is added with an upper limit on its total across all transfer nodes along each trip, model `WBTII`. To evaluate the benefits and contributions of our study, we compare the three versions of our model, with a conventional transfer synchronization approach in the literature, model `CM`, and the common practice of agencies in generating timetables, i.e., scheduling each line individually, model `NS`. Several performance metrics are defined to facilitate a thorough comparison between the models, such as the total and average transfer waiting times, average transfer buffer penalty per passenger, and total assigned transfer buffer times across the sub-network. In the second experiment, we explore the performance of our solution method for nine sub-networks with five, seven, and nine transfer nodes, which are drawn from the CTTN. We investigate how fast our solution method provides the best lower and upper bounds as well as the quality of the final solution compared to that obtained from a commercial solver. In the third experiment, as the main case study, we select from the CTTN a sub-network consisting of 12 transfer nodes. Since the sub-network has several transfer nodes involving lines with various headways, high- to low-frequency, we also explore the effect of incorporating headway regularity deviations on the solutions. To do so we solve the `WBTI` model with and without considering headway regularity deviations, `12NodesW` and `12Nodes`, respectively. In most previous studies, headway regularity has been only considered for departure times of buses from terminals. However, our model provides the opportunity to indirectly enforce headway regularity at all stops while considering different groups of passengers' waiting time: local, through, and transferring. In other words, our model quantifies the impacts of enforcing headway regularity on different groups of passengers. Analyzing the outputs of the above-mentioned experiments reveal the following insights and contributions of this study:

- The detailed structure of our transfer synchronization model, particularly the new successful transfer formulation, effectively improves the transferring waiting times. The solution of the `NB` model confirms that, even without the help of transfer buffer times, our formulation generates synchronized timetables with less total and average transfer waiting times compared to the models `CM` and `NS`.

- The solutions of the `WBTI` and `WBTII` models show the benefits of assigning transfer buffer times, specifically in terms of reducing transfer waiting time for transfers to low-frequency lines. Nevertheless, it is essential to take into account the induced penalty of extra in-vehicle time inflicted upon already on-board passengers. Our detailed formulation provides a quantitative trade-off between the advantages and disadvantages of assigning transfer buffer times. Moreover, our model considers the impact of adding transfer buffer times to bus dwell times on the arrival times of buses at their following stops, leading to more reliable bus trajectories and timetables.

- The computational burden does not only depend on the number of transfer nodes. In fact, several factors impact the number of iterations and computation times of our Lagrangian relaxation-based solution method. The factors include but are not limited to how the nodes are connected to one another, the sequence of transfer nodes for each line, and the headway combinations of intersecting lines played an important role. For instance, in one case with nine transfer nodes, the final solution with a 0.03% optimality gap is achieved in 65 minutes. However, the other nine-node network reached the final gap of 4.97% after 103 minutes.

- The results demonstrate the effectiveness of our proposed Lagrangian relaxation-based solution method. The performance of our solution method outweighs the commercial solver, GUROBI. In cases where our solution method solved nine-node and seven-node networks with 4.97% and 0.00% gap, respectively, GUROBI could not provide bounds during the same time budget.

- Evaluating headway regularity deviations for a 12-node network indicates the benefits of considering transfer buffer time and dwell time in timetabling models for effectively maintaining headway regularity. The solution shows that assigning transfer buffer time not only reduces transfer waiting times but also improves the regularity of bus arrival/departure times. In previous studies, headway regularity was only considered for departure times of buses from terminals. However, our model provides the opportunity to ensure headway regularity at all stops while considering the potential costs incurred by different groups of passengers: local, through, and transferring passengers. In other words, our model quantifies the trade-offs between the advantages and disadvantages of headway regularity on different groups of passengers. For instance, in some cases, with a low-frequency connecting line, the reduction in transfer waiting time with the aid of transfer buffer times might outweigh the cost of violating headway regularity.

The comparison between our model, a conventional transfer synchronization approach in the literature, and the common practice of agencies in generating timetables lends support to the detailed formulation proposed in this study. Our work contributes to a better understanding of the benefits of considering detailed successful transfers, bus dwell time determination, and bus capacity limit in timetabling models at the planning stage. In fact, in this study, we develop a base model formulation including essential details for stochastic and real-time transfer synchronization models. A solid baseline deterministic model, would indeed inform research efforts concerned with developing stochastic models of transfer synchronization. Our detailed model formulation can be generalized and used in stochastic transfer synchronization models as needed. Nevertheless, we identify some limitations in our study that we recommend investigating in future studies:

- Our model does not consider the potential impact of the produced timetables on demand levels and patterns, since we use fixed demand in our formulation. Future studies can explore the joint formulation of travel behaviour and timetabling models in an integrated framework.

- The inherent stochasticity in demand and travel times can heavily impact the performance of deployed timetables in the real world, in particular, when required details are not well treated in stochastic models. Therefore, future research should incorporate travel times and demand uncertainty, especially when bus dwell time determination and bus capacity limitations are incorporated in a model, to provide more realistic outputs.

- The given pre-defined network characteristics in our model can be considered as decision variables in future studies. For instance, it is beneficial to jointly design route frequencies and bus sizes in a transfer synchronization model.

- In this research, we considered trips in single directions, ignoring possible effects on layover times at terminals and assuming buses are always available to start their trips in the opposite direction according to the timetable. However, future research may consider trip cycle times and vehicle assignment to improve the applicability of transfer synchronization models in practice.

Funding. The authors would like to acknowledge the funding support by Trapeze Group, Ontario Centres of Excellence, Natural Sciences and Engineering Research Council, City of Toronto and York Regional Municipality.

References

Abdolmaleki, M., Masoud, N., Yin, Y., 2020. Transit timetable synchronization for transfer time minimization. Transportation Research Part B: Methodological 131, 143–159.

Ansari Esfeh, M., Wirasinghe, S., Saidi, S., Kattan, L., 2021. Waiting time and headway modelling for urban transit systems–a critical review and proposed approach. Transport Reviews 41, 141–163.

Ansarilari, Z., Nesheli, M.M., Srikukenthiran, S., Bodur, M., Shalaby, A., 2018. Multi-directional transfer time optimization at a single transfer node .

Ataeian, S., Solimanpur, M., Amiripour, M., Shankar, R., 2019. Synchronized timetables for bus rapid transit networks in small and large cities. Scientia Iranica .

Cao, Z., Ceder, A., Li, D., Zhang, S., 2019. Optimal synchronization and coordination of actual passenger-rail timetables. Journal of Intelligent Transportation Systems 23, 231–249.

Caprara, A., Fischetti, M., Toth, P., 2002. Modeling and solving the train timetabling problem. Operations research 50, 851–861.

Cats, O., Rufi, F.M., Koutsopoulos, H.N., 2014. Optimizing the number and location of time point stops. Public Transport 6, 215–235.

Cats, O., West, J., Eliasson, J., 2016. A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. Transportation Research Part B: Methodological 89, 43–57.

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

Ceder, A., Golany, B., Tal, O., 2001. Creating bus timetables with maximal synchronization. Transportation Research Part A: Policy and Practice 35, 913–928.

Chu, J.C., Korsesthakarn, K., Hsu, Y.T., Wu, H.Y., 2019. Models and a solution algorithm for planning transfer synchronization of bus timetables. Transportation Research Part E: Logistics and Transportation Review 131, 247–266.

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, 1202–1217.

Eranki, A., 2004. A model to create bus timetables to attain maximum synchronization considering waiting times at transfer stops .

Fouilhoux, P., Ibarra-Rojas, O.J., Kedad-Sidhoum, S., Rios-Solis, Y.A., 2016. Valid inequalities for the synchronization bus timetabling problem. European Journal of Operational Research 251, 442–450.

Gavriilidou, A., Cats, O., 2019. Reconciling transfer synchronization and service regularity: real-time control strategies using passenger data. Transportmetrica A: Transport Science 15, 215–243.

Gkiotsalitis, K., Eikenbroek, O.A., Cats, O., 2020. Robust bus scheduling considering transfer synchronizations, in: Transportation Research Board (TRB) 99th Annual Meeting.

Guo, X., Wu, J., Sun, H., Yang, X., Jin, J.G., Wang, D.Z., 2020. Scheduling synchronization in urban rail transit networks: Trade-offs between transfer passenger and last train operation. Transportation Research Part A: Policy and Practice 138, 463–490.

Gurobi Optimization LLC, 2020. Gurobi Optimizer reference manual.

Ibarra-Rojas, O.J., Giesen, R., Rios-Solis, Y.A., 2014. An integrated approach for timetabling and vehicle scheduling problems to analyze the trade-off between level of service and operating costs of transit networks. Transportation Research Part B: Methodological 70, 35–46.

Ibarra-Rojas, O.J., López-Irarragorri, F., Rios-Solis, Y.A., 2016. Multiperiod bus timetabling. Transportation Science 50, 805–822.

Ibarra-Rojas, O.J., Muñoz, J.C., Giesen, R., Knapp, P., 2019. Integrating frequency setting, timetabling, and route assignment to synchronize transit lines. Journal of Advanced Transportation 2019.

Ibarra-Rojas, O.J., Rios-Solis, Y.A., 2012. Synchronization of bus timetabling. Transportation Research Part B: Methodological 46, 599–614.

Liu, T., Ceder, A., 2016. Synchronization of public transport timetabling with multiple vehicle types. Transportation Research Record 2539, 84–93.

Liu, T., Ceder, A.A., 2018. Integrated public transport timetable synchronization and vehicle scheduling with demand assignment: A bi-objective bi-level model using deficit function approach. Transportation Research Part B: Methodological 117, 935–955.

Long, S., Meng, L., Miao, J., Hong, X., Corman, F., 2020. Synchronizing last trains of urban rail transit system to better serve passengers from late night trains of high-speed railway lines. Networks and Spatial Economics 20, 599–633.

Mazloumi, E., Mesbah, M., Ceder, A., Moridpour, S., Currie, G., 2012. Efficient transit schedule design of timing points: a comparison of ant colony and genetic algorithms. Transportation Research Part B: Methodological 46, 217–234.

Miller, E., Vaughan, J., Yusuf, B., Higuchi, S., 2018. Surface transit speed update report .

Parbo, J., Nielsen, O.A., Prato, C.G., 2014. User perspectives in public transport timetable optimisation. Transportation Research Part C: Emerging Technologies 48, 269–284.

Rapp, M.H., Gehner, C.D., 1967. Transfer optimization in an interactive graphic system for transit planning. Technical Report.

Salzborn, F., 1980. Scheduling bus systems with interchanges. Transportation Science 14, 211–231.

Schröder, M., Solchenbach, I., 2006. Optimization of transfer quality in regional public transit .

Shafahi, Y., Khani, A., 2010. A practical model for transfer optimization in a transit network: Model formulations and solutions. Transportation Research Part A: Policy and Practice 44, 377–389.

Silva-Soto, Y., Ibarra-Rojas, O., 2021. Timetabling with flexible frequencies to synchronise groups of bus lines at common stops. Transportmetrica A: Transport Science 17, 978–1001.

Srikukenthiran, S., Shalaby, A., 2017. Enabling large-scale transit microsimulation for disruption response support using the nexus platform. Public Transport 9, 411–435.

Tuzun Aksu, D., Yılmaz, S., 2014. Transit coordination with heterogeneous headways. Transportation Planning and Technology 37, 450–465.

Vera, J.R., Weintraub, A., Koenig, M., Bravo, G., Guignard, M., Barahona, F., 2003. A lagrangian relaxation approach for a machinery location problem in forest harvesting. Pesquisa Operacional 23, 111–128.

Voorhorst, J., Gkiotsalitis, K., 2020. Scheduling bus holding times in time horizons considering transfer synchronization and service regularity, in: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 1–6.

Wardman, M., 2004. Public transport values of time. Transport policy 11, 363–377.

Wirasinghe, S., 1980. Nearly optimal parameters for a rail/feeder-bus system on a rectangular grid. Transportation Research Part A: General 14, 33–40.

Wong, R.C., Yuen, T.W., Fung, K.W., Leung, J.M., 2008. Optimizing timetable synchronization for rail mass transit. Transportation Science 42, 57–69.

Wu, J., Liu, M., Sun, H., Li, T., Gao, Z., Wang, D.Z., 2015. Equity-based timetable synchronization optimization in urban subway network. Transportation Research Part C: Emerging Technologies 51, 1–18.

Wu, W., Liu, R., Jin, W., Ma, C., 2019. Stochastic bus schedule coordination considering demand assignment and rerouting of passengers. Transportation Research Part B: Methodological 121, 275–303.

Wu, Y., Yang, H., Tang, J., Yu, Y., 2016. Multi-objective re-synchronizing of bus timetable: Model, complexity and solution. Transportation Research Part C: Emerging Technologies 67, 149–168.

Yang, L., Zhou, X., 2014. Constraint reformulation and a lagrangian relaxation-based solution algorithm for a least expected time path problem. Transportation Research Part B: Methodological 59, 22–44.

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

Zhang, C., Gao, Y., Yang, L., Gao, Z., Qi, J., 2020. Joint optimization of train scheduling and maintenance planning in a railway network: A heuristic algorithm using lagrangian relaxation. Transportation Research Part B: Methodological 134, 64–92.

Zhigang, L., Jinsheng, S., Haixing, W., Wei, Y., 2007. Regional bus timetabling model with synchronization. Journal of Transportation Systems Engineering and Information Technology 7, 109–112.