



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/141348/>

Version: Accepted Version

Article:

Wu, W, Liu, R, Jin, W et al. (2019) Stochastic bus schedule coordination considering demand assignment and rerouting of passengers. *Transportation Research Part B: Methodological*, 121. pp. 275-303. ISSN: 0191-2615

<https://doi.org/10.1016/j.trb.2019.01.010>

© 2019 Published by Elsevier Ltd. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Please cite the paper as:

Wu, W., Liu, R., Jin, W. and Ma, C (2019) Stochastic bus schedule coordination considering demand assignment and rerouting of passengers. *Transportation Research Part B*. Accepted on 14 Jan 2019.

Stochastic bus schedule coordination considering demand assignment and rerouting of passengers

WeitiaoWu^a, Ronghui Liu^b, Wenzhou Jin^a and Changxi Ma^c

a. Department of Transportation Engineering, South China University of Technology, Guangzhou 510641,

China

b. Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, United Kingdom

c. School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

E-mail addresses: ctwtwu@scut.edu.cn; R.Liu@its.leeds.ac.uk; ctwzhjin@scut.edu.cn;

machangxi@mail.lzjtu.cn

Abstract:

Schedule coordination is a proven strategy to improve the connectivity and service quality for bus network, whereas current research mostly optimizes schedule design using the a priori knowledge of users' routings and ignores the behaviour reactions to coordination status. This study proposes a novel bus stochastic schedule coordination design with passenger rerouting in case of transfer failure. To this end, we develop a bi-level programming model in which the schedule design (headways and slack times) and passenger route choice are determined simultaneously via two travel strategies: non-adaptive and adaptive routings. In the second strategy, transfer passengers would modify their paths in case of missed connection. In this way, the expected network flow distribution is dependent on both the transfer reliability and network structure. The upper-level problem is formulated as a mixed integer non-linear program with the objective of minimizing the total system cost, including both the operating cost and user cost, while the lower-level problem is route choice (pre-trip and on-trip) model for timed-transfer service. A more generalized inter-ratio headways scenario is also taken into account. A heuristic algorithm and the method of successive averages are comprehensively applied for solving the bi-level model. Results show that when the rerouting behaviour is considered, more cost-effective schedule coordination scheme with less slack times can be achieved, and ignoring such effect would underestimate the efficacy of schedule coordination scheme.

Keywords: Public transport; Bus schedule coordination; Stochastic travel times; Passenger transfer; Rerouting

1. Introduction

In the era of internet-of-things, travellers expect seamless mobility solutions on the go and have access to up-to-the-minute traffic conditions and travel options. Seamless travel requires good public transport connectivity and cooperation among different lines. This can be done either before or after transit line alignment. Schedule coordination aims at adjusting the timetables or headways so as to make the vehicles on different lines arrive at the transfer station simultaneously as possible, which is a proven strategy after transit line alignment.

Although a well-planned synchronized scheme can considerably improve level of service by providing seamless transfer solutions, in practice, there exists large number of stochastic attributes in the public transport system: travel time, dwell time, adverse weather, demand, etc., which affects the materialization of the planned synchronized timetable (Yu et al., 2012; Wu, et al., 2015; Wu, et al., 2016). Such stochastic events cause arrival delay to transfer stations and missed connection for passengers. To counter such stochastic effect, in practice, transit agencies normally impose sufficient slack times into timetable to increase the probability of successful transfer coordination of connecting routes. However, the addition of slack times could bring two disbenefits: (1) from users' perspective, it will increase the on-board waiting time for those through passengers, and (2) from the viewpoint of operator, it increases the total round-trip time which in turn requires bigger fleet size to provide the same level of service. One of the greatest challenges to transit authorities is to improve service reliability and operation efficiency, which involves making a difficult trade-off between operation cost (by adding slack times) and the on-time arrival performance. Our study explores this idea.

It has been shown that larger headways (e.g., more than 10 min or even half an hour) make it easier for schedule coordination (e.g., Ting and Schonfeld, 2005; Wu et al., 2016), and that passengers perceives the value of waiting time twice as much as the in-vehicle travel time (Hollander and Liu, 2007). As a result, the extra waiting time costs caused by the transfer failure event will greatly change the relative utility between different routes, such that some passengers would modify their routes as response to the status of transfer coordination. With the provision of traffic information or route guidance along the trip, dynamic rerouting behaviour is prevalent under stochastic travel conditions (Zhao, et al, 2017; Xu, et al, 2017). As indicated in the empirical study by Kim et al. (2017), passengers would present lower stickiness on the same route with the increase in the number of alternative routes. The performance of transit operation is highly related to the temporal and spatial passenger flow distribution and running fleet (Liu and Singha, 2007; Sorrantina et al., 2006; Wu et al., 2017). The accurate transfer demand and transfer waiting time given to the operators are important determinants for the reservation of slack times and fleet management. As the first step towards an efficient transit system, understanding and modelling the interaction between travel behaviour and schedule coordination design can help the transit planners to reasonably allocate and schedule the available fleet in fine granularity. Since the amount of slack time is largely determined by transfer demand, one possible solution is to provide alternative path to transport passengers between the endpoints of the link affected by connection failures such that the stranded passengers are able to resume their journey. Therefore, besides imposing slacks, rerouting of passengers is a potential behaviour-driven strategy for delay management in

public transport.

Previous studies on bus schedule coordination generally assumed that behavioral pattern of passengers is independent of the synchronized schemes and connection status, which limits the service flexibility and operational efficiency. Distinct from prior research, considering passengers' rerouting behaviour, this paper develops an optimization framework for reliable timed transfer while achieving high efficiency that allows to react to delayed buses not only by slacks but also by rerouting of passengers. To this end, we first define and formulate the concept of rerouting in the context of timed transfer under connection uncertainty, where part of transfer passengers modify their paths in case of missed connection. We then solve the optimization of coordinated transfer by integrating such users' routing strategy. The switching rate is determined by not only the disutility of acceptable routes, but also the transfer failure rate, both of which are contingent on the service headways and slack times. Our findings show that the inclusion of rerouting considerably affects the transfer demand and network flow distribution, and more importantly, helps to reduce the unnecessary slack times, particularly under supply constraints. We thus suggest that the performance of schedule coordination could be improved in an ad-hoc manner, which provides new practical insights for fleet management.

The rest of the paper is organized as follows. In section 2, a review of previous works and our contributions are provided. In section 3, we formulate the model and solution algorithm. Section 4 performs experiments to verify the effectiveness of our model. Finally, Section 5 provides conclusion and future works.

2. Literature review and main contributions

Timetabling and scheduling problems for public transport have received considerable attention for many years due to their practical importance. Generally, schedule coordination can be carried out at both planning level and operation level. The former seeks to design a timetable to realize simultaneous arrivals at a transfer station as possible to reduce the transfer waiting time for passengers. The latter focuses on real-time disruption by managing unexpected delays to adhere operation to the pre-designed timetable as much as possible.

Ceder et al. (2001) developed a mixed integer linear programming model for the timetable generation problem, with the objective of maximizing the number of simultaneous vehicle arrivals at transfer hubs. Later, Eranki (2004) relaxed the synchronization condition proposed by Ceder et al. (2001), and allowed arrival time of buses from different lines at the connection nodes is not necessarily the same but within a targeted time window. In a more realistic setting, Ibarra-Rojas and Rios-Solis (2012) studied timetable synchronization problem for a common-line situation where different bus lines share the same route segment, in which bus bunching avoidance is desirable. Later, Ibarra-Rojas et al (2015) further expanded the model by considering different planning periods for different lines. The common objective of Ibarra-Rojas and colleagues' work is to improve passenger transfers and reduce bus bunching. In contrast, Wong et al (2008) established a mixed integer programming method for non-periodic timetable synchronization problem, which is realized by jointly adjusting travel time, dwell time, dispatching time and headways. Recently, Liu and Ceder (2017) integrated the timetable synchronization and vehicle scheduling using deficit function.

In practice, buses usually operate in a large open-ended environment, which entails large amount of uncertainty that prevents vehicles from operating according to schedule and hinders fully realization of the efficiency of schedule coordination. As a result, it is difficult to maintain vehicles consistent with the planned schedule determined by deterministic models in uncertain environment. In order to alleviate the uncertainty, many studies have focused on inserting control strategies into schedule, either at the planning or operation stage.

At the planning level, the most commonly used strategy is to impose slack times in the timetable. Ting and Schonfeld (2005) developed a model of schedule coordination by optimization of headways and slack times, with the objective of minimizing the net total costs. Chowdhury and Chien (2011) studied the schedule coordination problem in an intermodal transfer station in New Jersey Coast Line, where passenger transfer between train and bus as well as between buses. The design variables to be optimized include vehicle size, headways and slack times. Thereafter, Wu et al. (2015) investigated a timetabling model by adding slack times onto mitigate the travel time variability. They built upon the original work of Ceder et al. (2001) and developed a stochastic integer programming model where decision variables consist of slack times and departure times.

In parallel with studies at the planning level, there also have been a few studies at the operation level which attempt to impose real-time control strategies into schedule in a real-time manner, such as holding, stop-skip and speed control. As summarized in a review by Ibarra-Rojas et al. (2015), holding is an adaptive decision method and is the most commonly used dynamic control strategy that can be implemented at both intermediate bus stops and at transfer nodes. The goal of the former is to reduce bus bunching and regular headways along a bus corridor, while the latter is to allow well-timed passenger transfers. Ting and Schonfeld (2007) developed a dynamic holding control model for a transfer station. Chuang and Shalaby (2007) applied holding control to realize connection protection, and evaluated it at an intermodal transfer station in the city of Brampton, Canada. Some researchers also applied the holding control to deal with stochastic disturbances or delayed arrivals at transfer nodes (e.g. Dessouky et al., 1999, 2003; Chowdhury and Chien, 2001; Yu et al., 2012). Daganzo and Pilachowski (2011) proposed an adaptive speed control strategy in a two-way looking manner, which is based on the spacing both in the front and rear of each bus. Liu et al (2014) proposed an inter-vehicle communication scheme to optimize the schedule coordination, where two operational tactics: speed control and holding at transfer point were employed by using real-time information. Hadas and Ceder (2010) developed simulation models for transfer synchronization through optimal combinations of the selected tactics, such as holding, stop-skipping and short-turning. Nesheli and Ceder (2014) refined the simulation framework by the introduction of skip-segment strategy apart from skipping a stop. Later, Wu et al. (2016) put forwards a new strategy named “safety control margin” to improve the operation efficiency and transfer reliability of schedule coordination, by integrating dynamic holding control and slacks at both planning and operation levels.

The performance of schedule coordination scheme largely depends on the interaction between transport supply and the demand side in the bus network. However, most of existing studies on schedule coordination assume passengers’ route choice being independent of the timetable, and they choose paths pursuing the

minimal cost without considering the transfer connection failure (e.g., Ting and Schonfeld, 2005; Wu et al, 2016). In other words, passenger routes are predetermined in the optimization process given the planned timetable. This assumption is difficult to defend since in reality the passengers' route choice could change with synchronized timetable and transfer reliability. Recent works for timetable synchronization considering the interaction between timetabling and passengers' route choice are Parto et al (2014) and Liu and Ceder (2017), assuming that passengers follow a fixed route through the network with given timetable (i.e., static user equilibrium). Their models are deterministic in that the stochastic travel time and the resulting connection failure has not been considered. However, in the presence of missed connection and rerouting decisions along the journey, the passenger flow may not evolve to an equilibrium state. To increase the operational accuracy and behavioural realism, it is necessary to capture both the pre-trip and on-trip decisions in the context of schedule coordination. Allowing such realistic passenger behaviour presents difficulties to the optimization of schedule coordination. The main challenge is how to formulate jointly the departure route choice, adaptive on-trip rerouting options and synchronized operation design.

In this study, we develop a more generalized schedule coordination model into which the transfer reliability and the respective behaviour reactions (i.e., rerouting) are explicitly incorporated. The transfer reliability refers to the expected probability of successful connection. The reservation of slack times at transfer hub could significantly increase the success of transfer coordination, the optimization is performed with the view of exploring the optimal headways and slack times to reduce the systematic total costs, including user cost and operator cost.

The contributions of the paper are: (1) we propose a unified framework for modelling route choice (pre-trip and on-trip) in the context of schedule coordination, where the transfer passengers may modify their originally planned path in the event of missed connection. The model is scalable allowing passengers to modify the path choice multiple times; (2) we build the interdependencies between uncertainty of bus delay arrival times, the transfer failure rate and rerouting probability, in which way the interaction between the schedule design and route choice is handled; (3) we demonstrate that by including rerouting behaviour in the travel strategy, more cost-effective schedule plan with less slack times can be materialized. *To the best of our knowledge, this is the first attempt in optimizing bus schedule coordination scheme by integrating transfer reliability and adaptive rerouting behaviour from the systematic point of view.*

3. Model formulation

3.1. Problem description

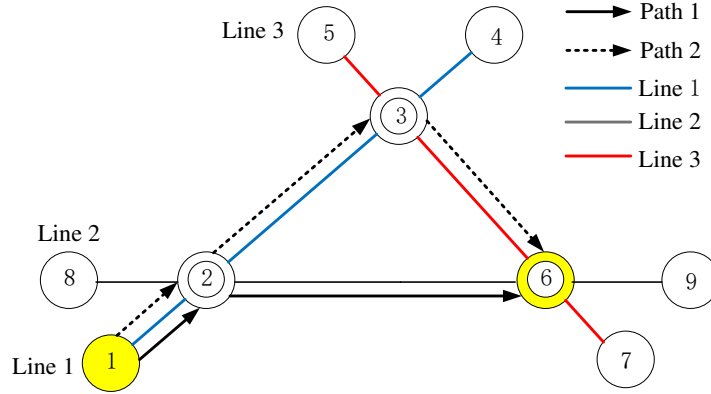


Fig.1 A diagrammatic illustration of passenger rerouting under missed connection in a looped network

This paper seeks to develop an enhanced stochastic bus schedule coordination model, more specifically by modelling the pre-trip and on-trip route choice, and investigates how the passenger route choice interact with different operational settings. To achieve so, we firstly present a bus schedule coordination model with static route choice without rerouting behaviour, then extend the basic model by explicitly modelling passenger rerouting as another way of handling uncertainty. The models are compared under different given demands and supply constraints.

Consider a bus network with time-transfer where a fixed number of passengers want to go from node 1 to node 6 (see Fig.1 for illustration). There are two possible itineraries along paths 1 and 2, in which path 1 is the shorter path under the condition of successful coordination. For simplicity here we assume that passengers take the shortest route to a destination. If the possibility of passengers rerouting is not considered, the number of passengers affected by missed connection is the total demand travelling from node 1 to node 6, which could lead to unnecessary long slack time for node 2. However, if the passengers is allowed to adapt their path to the prevailing delay situation, it turns out that the negative effect of missing connection can be partially mitigated. There are two options in this case: some passengers would stay at node 2 and wait for the next bus on line 2, while the other group may switch to path 2. In other words, they will stay at the delayed bus until arriving at node 3 and then reach the destination by transferring to line 3.

The example shows that the actual demand affected by missed connection is only part of the total demand travelling from node 1 to 6. As a result, the detrimental effects of missed connection with rerouting of passengers is smaller than those without for the same amount of slack time. In other words, the optimal allocated slack times with rerouting behaviour may be smaller than those without rerouting.

To this end, we define the concept of rerouting behaviour in the context of schedule coordination under connection uncertainty as follows: Passengers determine their route beforehand (via checking the route and timetables). In case of missed connection, passengers are assumed to compare travel cost to that of the currently used path in deciding whether to switch to another alternative. In this way, both pre-trip and on-trip decisions have been handled in the bus schedule planning.

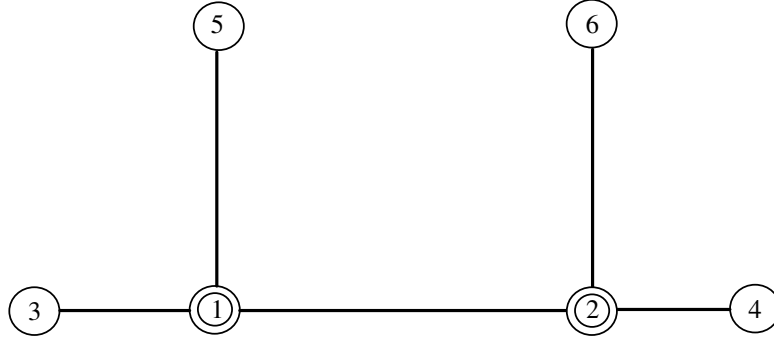


Fig.2 A diagrammatic illustration of trunk-and-feeder network

Note that whether there exists another alternative path depends on the bus network structure, while the swapping demand is determined by utility differences between the originally planned path and alternative paths. Generally, bus networks present different topological structures, such as trunk-and-feeder (Sivakumaran, K., et al, 2012; Gschwender, A., et al, 2016), looped (Ting and Schonfeld, 2005; Wu et al, 2016), and ring-radial structure (Chen, et al, 2015; Saidi, S, et al, 2016). For trunk-and-feeder network, as shown in Fig.2, any OD pair between two branches should run through the main ‘trunk’ route, thus no alternative could be provided by the spatial network. Such a network structure is however beyond the scope of this study.

3.2. Network representation, assumption and notations

Suppose a general transit network consisting of a set of bus lines and stations (node) where passengers can board, alight or transfer. For illustrative purpose, we adopt the network in Huang et al (2016) as an example. In this network (Fig.3), a bus line is a fixed path connecting two terminals, for instance, bus line L_1 can be seen to run between terminals N_1 and N_4 . A transit link is a segment of a bus line connecting two consecutive stations, represented as $N_1 \xrightarrow{L_1} N_2$. A *section-based path* is represented by a sequence of nodes and links used on this trip, e.g., $N_1 \xrightarrow{L_1} N_2 \xrightarrow{L_3} N_4$. Due to the consolidation of bus lines in a route section, a section-based path can be decomposed into several *line-based paths*. For example, the section-based path $N_1 \xrightarrow{S_1(L_1, L_3)} N_4$ can be extended to two line-based paths: $N_1 \xrightarrow{L_1} N_4$ and $N_1 \xrightarrow{L_3} N_4$.

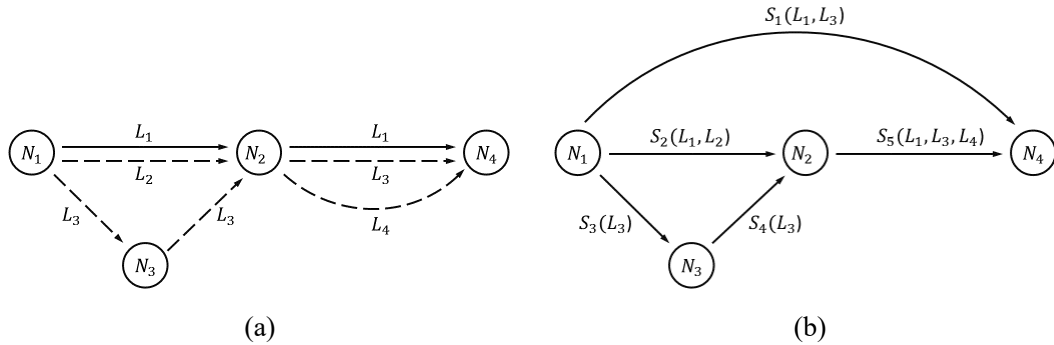


Fig.3 Transit network representation using: (a) lines and (b) route sections (adapted from Huang et al., 2016)

Without loss of generality, the assumptions considered in developing the model are the following:

(A1) Route information, including passenger demand and bus travel time, are known and fixed in the planning period.

(A2) Since the minimal passengers' walking time at the transfer station can be simply added onto the vehicle running time as input, it is reasonable to assume that the transfer time between buses at a transfer station is zero.

(A3) The model focuses on the schedule design over a specific planning horizon in which passenger demand and travel time retain their characteristics. However, the model can be extended to incorporate demand and running time dynamics in multi-period schedule coordination planning to capture the time-dependent operating conditions, which has been left for future research.

The notations adopted in this paper are given in Table 1.

Table 1 Primary notations used in the article

Notations	Definitions
Indices	
k	the bus line index
s	the index of route section
m	indices of transfer nodes
p	the index of transit path
Model parameters	
B	unit vehicle operation cost
μ_w	value of passenger waiting time and delayed connection time
μ_c	value of missed connection time
t_s	average travel time on link s
Ca	vehicle capacity
$f(d_{mk})$	the probability density function of delayed arrival at transfer node m on line k
θ	the degree of passengers' perception of the path travel time
Auxiliary variables	
Ca_s	the capacity of link s
x_s	the flow of link s
T_s	perceived in-vehicle travel time on link s
Q_s	boarding demand on link s
Q_{mk}	through demand at transfer node m on line k
$Q_m^{kk'}$	transfer demand at transfer node m between line k and its connecting line k' without rerouting; or the transfer demand for missed connection with rerouting
$Qt_m^{kk'}$	transfer demand for delay connection at transfer node m between line k and its connecting line k' with rerouting
f_p^w	section-based flow of path p between OD pair w without rerouting
\hat{f}_p^w	line-based flow of path p between OD pair w without rerouting
\hat{f}_{pn}^w	non-adaptive flow of path p between OD pair w under n^{th} transfer
\tilde{f}_{pn}^w	adaptive flow detouring through alternative path \bar{p} between OD pair w under n^{th} transfer

\vec{f}_{pn}^w	the expected incoming flow on path p under n^{th} transfer
Pr_{pj}^w	switching rate at j^{th} transfer station on path p between OD pair w
$P_m^{kk'}$	transfer failure rate between line k and k' at transfer node m
Pc_{pj}^w	transfer failure rate at j^{th} transfer station on path p between OD pair w
ξ_{sp}^w	1 if link s is the first section of path p between OD pair w , and 0 otherwise
$\tau_m^{kk'}$	1 if line k and line k' share transfer node m , and 0 otherwise
δ_{sp}^w	1 if link s is on path p between OD pair w , and 0 otherwise
b_{kp}	1 if line k is on path p , and 0 otherwise
φ_k^m	1 if node m (not the terminal) is on line k , and 0 otherwise.
ψ_{pj}^m	1 if node m is the j^{th} transfer station on path p between OD pair w , and 0 otherwise.
$Z_m^{kk'}$	inter-cycle transfer waiting time from line k to line k' at node m
U_p^w	the expected travel disutility on path p between OD pair w
\tilde{T}_m^p	the expected transfer waiting time cost at transfer node m on path p .
C_o	operation cost
C_I	in-vehicle travel time cost
C_w	initial waiting time cost
C_c	missed connection cost
C_f	delayed connection cost
C_g	inter-cycle delay cost
C	system total cost
Decision variables	
s_{mk}	slack time on line k at transfer node m
h_k	departure headway on line k

3.3. Individual cost formulation without rerouting

Without loss of generality, we assume the delayed arrival time follow a given distribution. To assure a high probability of scheduled transfer at the planning level, slack times are often added into the schedule to mitigate the effect of delayed arrival time. The problem of stochastic bus schedule coordination design is to determine the headways and slack times that corresponds to the minimum value of total costs, which involves the trade-off between various cost components. In order to operate the schedule periodically, the headways of different bus lines should be set as common or inter-ratio. This leads to the following two scenarios:

- (a) *common headway (CH)* where the headways of different bus lines are identical, i.e., $h_k = h$, where h is the common headway.
- (b) *inter-ratio headway (IR)* where the optimized headways of different lines are integer multiples of a base cycle, i.e., $h_k = \beta_k y$, where y is the base cycle and β_k is a positive integer number. If $\beta_k = 1$, the integer-ratio headway scenario is equivalent to common headway. In other words, the common headway scenario is a special case of inter-ratio headway.

For coordinated operations, the system costs are determined at the planning level including: initial waiting time cost C_w , in-vehicle travel time cost C_I , operating cost C_o , and the induced slack cost C_s cost caused by the introduction of slack times. Moreover, the transfer waiting costs could be separated into two (or three

for IR scenarios) components: (a) missed connection cost C_c ; (b) delayed connection cost C_f ; and (c) for IR operation only, the inter-cycle transfer waiting time cost C_g . We start from extending the schedule coordination design model of Wu et al (2016) to include passenger pre-trip and en-route route choices.

3.3.1 Induced slack time cost

The reservation of slack times would induce additional costs for both passengers and operator. From the perspective of passengers, it increases the on-board travel time for through passengers; while from perspective of operator, it increases the roundtrip time and resulting fleet size. As a result, the induced slack time cost takes the following forms:

$$C_s = \sum_m \sum_k \left(\mu_w Q_{mk} + \frac{B}{h_k} \right) s_{mk} \quad (1)$$

where the second term represents the product of extra fleet size s_{mk}/h_k and the unit vehicle operating cost. Q_{mk} is the number of passenger passing through transfer station m on bus line k and can be computed by:

$$Q_{mk} = \sum_w \sum_p \dot{f}_p^w \varphi_k^m b_{kp} \quad (2)$$

where b_{kp} equals 1 if bus line k is included in path p , and 0 otherwise. $\varphi_k^m = 1$ if node m (not the terminal) is on line k , and 0 otherwise.

Note that in Eq.(2) the path flow here is line-specified, which can be extracted from section-based path flow. The number of passengers using each line in a given route section is proportional to its relative aggregated frequency (De Cea and Fernández, 1993), thus passenger flow on route section s that uses bus line k can be expressed as:

$$x_s^k = \pi_s^k x_s, \quad k \in A_s \quad (3)$$

In this spirit, for a transfer path consisting of a sequence of bus lines k , the line-based path flow takes the following iterative forms:

$$\dot{f}_p^w = \prod_k \pi_s^k \dot{f}_p^w, \quad k, s \in p \quad (4)$$

Note that the direct path, where transfer is neglected, is a special case of a transfer path where the number of bus lines used by travelers is only one, since there is no other route to be selected as all of the demand will travel through route section s which coincides with pair w .

3.3.2 Missed connection cost at transfer stations

Missed connection event occurs when the connecting bus leaves before the transfer passengers arrive. The missed connection cost is the summation of product of transfer demand and the expected missed connection time, which is expressed as follows:

$$C_c = \mu_c \sum_m \sum_k \sum_{k'} Q_m^{kk'} T_c^{kk'} \quad (5)$$

where $T_c^{kk'}$ represents the expected missed connection time depending on the delayed distribution and slack times, of which the detailed derivation is provided in Appendix 1.

Similar to the through passenger demand, the calculation of transfer demand concerns with the line-

based path flow. For the case without rerouting behaviour (passengers select a path a priori and continue along it throughout their journey), the transfer flow at transfer station m between line k and line k' can be simply expressed as follows:

$$Q_m^{kk'} = \sum_w \sum_p \hat{f}_p^w b_{kp} b_{k'p} \tau_m^{kk'} \quad (6)$$

where $\tau_m^{kk'} = 1$ if and only if line k and line k' share transfer node m . Otherwise, $\tau_m^{kk'} = 0$.

3.3.3 Delayed connection cost at transfer stations

Unlike missed connection case, transfer passengers could make a successful connection by waiting for relatively shorter time instead of a full headway. The total delayed connection cost is associated with the product of transfer demand and the expected total delayed connection time. Note that for the case without rerouting, the formulation of transfer demand is identical to that of missed connection as in Eq.(6).

$$C_f = \mu_w \sum_m \sum_k \sum_{k'} Q_m^{kk'} T_f^{kk'} \quad (7)$$

where $T_f^{kk'}$ represents the expected delayed connection time (see Appendix 2 for deviation).

3.3.4 Transfer waiting time under integer-ratio headways

In IR scenarios, the average transfer demand between two bus lines is related to the integer ratio of two headways. This affects the missed connection cost C_c and the delayed connection cost C_f . Following Ting and Schonfeld (2005), since vehicles of two lines encounter at a transfer node every time interval $h_k h_{k'} / g_{kk'} \gamma$, the average transfer demand is $Q_m^{kk'} g_{kk'} \gamma / h_{k'}$, where $g_{kk'}$ stands for the greatest common divisor of β_k and $\beta_{k'}$, i.e., $g_{kk'} = gcd(\beta_k, \beta_{k'})$. Therefore, the missed connection cost and delayed connection cost for IR follow the same formulations as those for the CH scenarios in Eqs.(6) and (7), but to replace $Q_m^{kk'}$ by $Q_m^{kk'} g_{kk'} \gamma / h_{k'}$.

The formulations of induced slack time cost C_s for IR scenarios are the same as their corresponding CH operations as in Eq. (1). This is because the formulations are not related to the transfer demand.

Unlike the CH scenario, the inter-cycle transfer waiting cost C_g should be included in the transfer cost in IR scenario. The inter-cycle waiting cost C_g can be presented as follows:

$$C_g = \mu_w \sum_k \sum_{k'} \sum_m Q_m^{kk'} Z_m^{kk'} \quad (8)$$

$$Z_m^{kk'} = \frac{h_{k'}}{2} + s_{mk'} - \frac{1}{2} g_{kk'} \gamma \quad (9)$$

Eq. (9) represents the total transfer waiting time from line k to its connecting line k' , where γ is the base cycle. The term $h_{k'}/2 + s_{mk'}$ represents the upper bound of the average waiting time for line k' . The second term $g_{kk'} \gamma / 2$ may be interpreted as the transfer waiting time reduction due to different line headways. If the common headway is selected as the solution, the first term is eliminated and the waiting time equals to the slack time of that line.

3.3.5 In-vehicle travel time cost

There may exist common-line corridors in the bus network (Schmöcker et al., 2016), the link-specific travel time is associated with the line-specific travel times and relative frequency. let t_s^k be the in-vehicle travel time for bus line k on route section s . Then, the expected in-vehicle travel time t_s on route section s is equal to the weighted average of t_s^k of all lines on this route section. That is,

$$t_s = \sum_{k \in A_s} \pi_s^k t_s^k \quad (10)$$

where A_s is the set of attractive lines on common link s . π_s^k is the relative frequency defined as follows:

$$\pi_s^k = \frac{1/h_k}{1/h_s}, \quad \forall k \in A_s \quad (11)$$

where the joint headway for the common section is the of service frequency of all attractive lines, i.e., $h_s = 1/(\sum_{k \in A_s} (1/h_k))$.

To capture the passenger's discomfort effect, the perceived in-vehicle travel time is introduced which depends on the link flow x_s and corresponding travel time t_s , which is given by:

$$T_s = \begin{cases} t_s \left[1 + a \left(\frac{x_s}{Ca_s} \right)^b \right] & \text{for } x_s > Ca_s \\ t_s & \text{otherwise} \end{cases} \quad (12)$$

where Ca_s represent the link capacity; a and b are the model parameters. Eq.(12) resembles the Bureau of Public Roads (BPR) function in relation to both travel time and the level of congestion. Note that the congestion cost of interest here doesn't include the extra delay induced by the leftover passengers.

The passenger flow on link s is related to the path flow in the transfer network, and can be estimated by:

$$x_s = \sum_w \sum_p f_p^w \delta_{sp}^w \quad (13)$$

where δ_{sp}^w is the indicator, which equals 1 if route section s is on path p between OD pair w .

In Eq.(12), the capacity on link s can be estimated by the product of vehicle capacity and service frequency (the reciprocal of headway), i.e.,

$$Ca_s = \gamma Ca/h_s \quad (14)$$

where Ca represents the vehicle capacity. $\gamma = 60$ min/h if the unit for headway is minutes and the line capacity is passengers per hour.

As a result, the total perceived in-vehicle time cost can be calculated as the product of the link flow and the corresponding perceived in-vehicle travel time and expressed as follows:

$$C_I = \mu_w \sum_s x_s T_s \quad (15)$$

3.3.6 Initial waiting time cost

The initial boarding demand includes the passengers originating at intermediate stops and transfer stations. The initial waiting time is associated with the departure headways. In the presence of common lines, passengers who boarding at the common station can travel by any of the lines. For random passenger arrivals, the expected initial waiting time associated with link s can be calculated as half of the expected joint headway:

$$EW_s = \frac{1}{2}E(h_s) \quad (16)$$

where $E(h_s)$ is the expected joint headway for the common link and can be estimated by the design headway, that is, $E(h_s) = h_s$. As the headway becomes larger, passengers are likely to time their arrivals with the prescribed arrival/departure times. This can be handled by simply discounting a factor as in Mpccia and Laporte (2016), so there is no conceptual difficulty in modelling more complex behaviour.

The total initial waiting time cost is related to the expected initial waiting time and the boarding demand. The waiting time on link s catering to a total boarding demand Q_s is $Q_s \cdot EW_s$. Summing up the link-specific waiting time yields the total waiting time as follows:

$$C_w = \mu_w \sum_s Q_s EW_s \quad (17)$$

In the context of schedule coordination, passengers only have to wait for the arrived vehicle at the first section of their path, and they could make a relatively smooth transfer in the sequence of their path. The corresponding boarding demand on link s is:

$$Q_s = \sum_w \sum_p f_p^w \xi_{sp}^w \quad (18)$$

where f_p^w is flow on path p between OD pair w without rerouting. Note that $\xi_{sp}^w = 1$ if link s is the first section of path p between OD pair w . Otherwise, $\xi_{sp}^w = 0$.

3.3.7 Total system cost

Following Wu et al. (2016), the line-specific operation cost is the product of required fleet size T_k/h_k on line k and the unit operation cost. Summing up the line-specific operation cost yields the total operation cost as:

$$C_o = \sum_k \frac{BT_k}{h_k} \quad (19)$$

Then system total cost is the summation of user cost and operation cost, that is:

$$C = C_s + C_c + C_f + C_g + C_I + C_w + C_o \quad (20)$$

3.4. Cost formulation for coordinated operation with rerouting

3.4.1 Model of passenger rerouting behaviour

In this section, we introduce passenger rerouting behaviour, where transfer passengers modify their route in the event of missed connection. We analyze the effect of rerouting behaviour on the network flow presentation, more specifically the formulations of transfer and through demand. The route choice decision during the entire journey could be divided into two phases, namely, pre-trip and on-trip. A customary approach to describe the passenger route choice is user equilibrium or transit assignment (e.g., Yu, et al, 2015; Parto et al, 2014; Szeto and Jiang, 2014). However, after an implementation of bus schedule coordination scheme, the network flow could not achieve an equilibrium state in the presence of inherent transfer unreliability and the resulting rerouting behaviour, hence it is problematic to take the equilibrium-based indexes for demand representation in the context of schedule coordination with connection uncertainty. As such, the results of user equilibrium or transit assignment models only represent the pre-trip route choice.

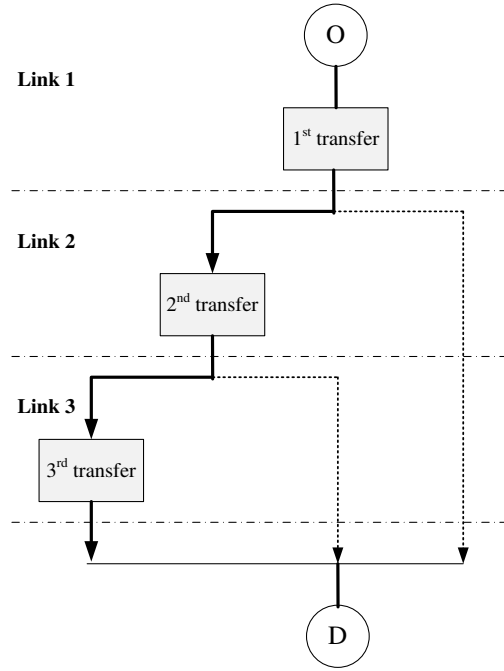


Fig.4 Illustration of stochastic rerouting process

Whether a passenger will keep her originally planned path once encountering a missed connection is a stochastic process. Assuming that a passenger would not modify their transfers for more than twice, the route-searching process for a passenger traveling is illustrated in Fig.1. The solid line represents the originally planned path made at the pre-trip stage, while the dashed lines represents the en-route alternative paths. The stochastic effect that the given passenger modifies the path when she encounters a transfer can be represented by a probability. If the given passenger modifies the path, she keeps this path until reaching the destination. The individual route-searching process can be regarded as a stochastic process. In the current problem, consider the trip chain of a given passenger, the state variable is whether the given passenger is stick to the originally planned path, while the initial state of the given passenger is the originally planned path at the pre-trip state. The probability of a state is related to the transfer failure rate and switching rate as will be derived below.

The microscopic stochastic rerouting behaviour is likely to have an impact on the network flow distribution, and requires modifications when applying traditional models. To capture the effect of travellers' decision along the trip chains on the network flow distribution, a two-phase algorithm is proposed to convert the individual stochastic process into the aggregate flow swaps. The idea is to partition the route choice problem into flow generation and redistribution sub-problems that are solved sequentially, where the results for one are data inputs for the next. In the first phase, all users in the system make pre-trip route choice decisions and generate the original section-based path flow f_p^w and line-based path flow \hat{f}_p^w , which could be solved by the transit assignment sub-model (see later in Section 4). In the second phase, a route choice option is presented to passengers which can lead to flow swaps. Accordingly, the original network flows should to be redistributed. Fig.5 illustrates the unified framework for how original network flow process across different model components, in which flow conservation should be ensured in the redistribution.

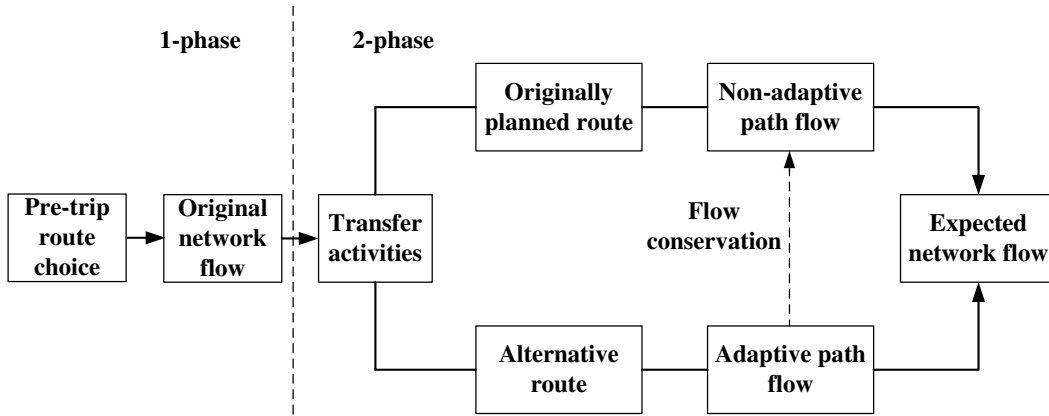


Fig.5 Flow redistribution framework

Given the original path flow, the adaptive path flow (rerouting demand) can be obtained by the utility differences between attractive routes. Subsequently, with the principle of flow conservation, the adaptive flow can be calculated based on the original path flow minus non-adaptive path flow. As a result, the actual path flow and resulting expected network flows are also known by adding up the non-adaptive and adaptive flow for overlapping sections.

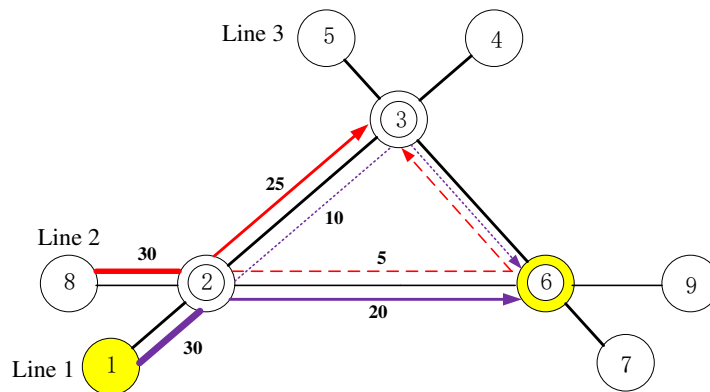


Fig.6 Illustration of flow redistribution

To illustrate the flow redistribution, we show in Fig.6, suppose there are two OD pairs: 1→6 and 8→3, each of which has an OD demand of 30. Their dominant (or shortest) paths are denoted by solid lines. Let us assume that adaptive path flows are 5 and 10, respectively (as shown by dashed lines). On adding up the non-adaptive flow and adaptive flow on specific links, the expected link flows for (2, 3) and (2, 6) are 35 (= 25+10) and 25 (= 20+5). As the original link flows are both 30, we can conclude that the redistributed link flows could be either smaller or larger than the original ones. As a result, the network-wide transfer demand and through demand should be reorganized accordingly. Since the preceding travel path and following alternative path may share the same bus line at a transfer station, the rerouting behaviour would lead to an increase of through demand at the expense of reduced transfer demand, of which the effect on the schedule coordination design will be investigated in Section 4.

3.4.2 Transfer failure rate

The higher missed connection probability, the greater the possibility of passengers reroute. In order to

calculate the flow swaps at a transfer node, the concept of transfer failure rate is given, which refers to the probability of events where at least one passenger misses the scheduled transfer bus and have to wait for the next bus.

By definition, the transfer failure rate rests with the delayed arrival distribution, and how service headways and slack times are configured. More specifically, the transfer failure rate $P_m^{kk'}$ at specific transfer station can be provided as follows, and the detailed derivation is provided in Appendix 3.

$$P_m^{kk'} = \int_{s_{mk}}^{h_k} \int_0^{s_{mk'}} f(d_{mk})f(d_{mk'}) dd_{mk}dd_{mk'} + \int_{s_{mk}}^{h_k} \int_{s_{mk'}}^{d_{mk}-s_{mk}+s_{mk'}} f(d_{mk})f(d_{mk'}) dd_{mk}dd_{mk'} \quad (21)$$

3.4.3 Transfer and through demand after flow redistribution

As implied previously, passenger rerouting would change transfer demand and through demand at the transfer hub. Therefore, compared to the model without rerouting, the system costs with rerouting contains four different components: induced slack time cost, the missed connection cost, delayed connection cost and transfer waiting time cost under integer-ratio headways.

Now we discuss the formulations of transfer and through demand after flow redistribution, provided that the original line-based path flow \hat{f}_p^w is available. To this end, we introduce two new set of variables (\hat{f}_{pn}^w , $\tilde{f}_{\bar{p}n}^w$) to record the number of passengers who stick to path p and use the alternative path \bar{p} under n^{th} transfer, respectively. The idea behind this is that the original path flow would be routed into different acceptable paths (and thus discounted) once undergoing one transfer.

Modifying Eq.(6) gives the following transfer demand for missed connection with rerouting behaviour as:

$$Q_m^{kk'} = \sum_w \sum_p \sum_n \hat{f}_{pn}^w b_{kp} b_{k'p} \tau_m^{kk'} + \sum_w \sum_{\bar{p}} \sum_n \tilde{f}_{\bar{p}n}^w b_{k\bar{p}} b_{k'\bar{p}} \tau_m^{kk'} \quad (22)$$

where the first term denotes the non-adaptive flow, and the second term denotes the adaptive flow from other paths. Note that both path p and \bar{p} traverse between line k and line k' . The underlying assumption is that the group of swapping passengers would update their route once at most.

The swapping demand stems from two determinants: (A) the frequency of connection failures, namely, the higher failure frequency, the higher likelihood of passenger rerouting. This is because each rerouting decision is triggered by missed connection; (B) the switching rate under connection failure, which is calculated as the disutility difference between the acceptable paths (As one may see later in this section). Since passengers may complete their travel using multiple bus lines, the original path flow should be discounted once undergoing a transfer station.

Proposition 1. Given the transfer failure rate and switching rate at the designated stop, the expected non-adaptive flow of path p between OD pair w under n^{th} transfer can be computed as follows:

$$\hat{f}_{pn}^w = \hat{f}_p^w \prod_{j=1}^n (1 - Pr_{pj}^w Pc_{pj}^w)$$

where Pr_{pj}^w and Pc_{pj}^w are the switching rate and transfer failure rate at j^{th} transfer station on path p between OD pair w , respectively.

Proof. Since rerouting behaviour only occurs in case of missed connection, the expected rerouting flow is

given by the original flow multiplied with the respective transfer failure rate and switching rate. With the conservation of flows, the expected non-adaptive flow for path p between OD pair w after undergoing the first transfer is the difference between the original (pre-trip) flow and the rerouting flow, that is,

$$\hat{f}_{p1}^w = \dot{f}_p^w - \dot{f}_p^w Pr_{p1}^w Pc_{p1}^w = \dot{f}_p^w (1 - Pr_{p1}^w Pc_{p1}^w)$$

Similarly, the expected non-adaptive flow after undergoing the second transfer can be calculated as follows:

$$\hat{f}_{p2}^w = \hat{f}_{p1}^w - \hat{f}_{p1}^w Pr_{p2}^w Pc_{p2}^w = \dot{f}_p^w (1 - Pr_{p1}^w Pc_{p1}^w) (1 - Pr_{p2}^w Pc_{p2}^w)$$

In this vein, it is not difficult to conclude that the expected non-adaptive flow of path p under n^{th} transfer could take the following iterative forms:

$$\hat{f}_{pn}^w = \dot{f}_p^w \prod_{j=1}^n (1 - Pr_{pj}^w Pc_{pj}^w)$$

Q.E.D.

Proposition 2. Based on the routing demand conservation constraints, the expected adaptive flow of the alternative path \bar{p} under n^{th} transfer is given as follows:

$$\tilde{f}_{\bar{p}n}^w = \begin{cases} \dot{f}_p^w Pr_{\bar{p}n}^w Pc_{\bar{p}n}^w, & \text{for } n = 1 \\ \dot{f}_p^w \prod_{j=1}^{n-1} (1 - Pr_{pj}^w Pc_{pj}^w) Pr_{\bar{p}n}^w Pc_{\bar{p}n}^w, & \text{for } n > 1 \end{cases}$$

Proof. When $n = 1$, the expected adaptive flow of the alternative path \bar{p} is simply the original flow multiplied by the transfer failure rate and switching rate, i.e.,

$$\tilde{f}_{\bar{p}1}^w = \dot{f}_p^w Pr_{\bar{p}1}^w Pc_{\bar{p}1}^w$$

When $n > 1$, according to Proposition 1, the expected incoming flow of path p before undergoing $(n - 1)^{\text{th}}$ transfer is

$$\hat{f}_{p,n-1}^w = \dot{f}_p^w \prod_{j=1}^{n-1} (1 - Pr_{pj}^w Pc_{pj}^w)$$

The conservation of flows requires that the rerouting flow (to alternative path) is the difference of incoming flows between $(n - 1)^{\text{th}}$ and n^{th} transfer on original path, that is,

$$\tilde{f}_{\bar{p}n}^w = \hat{f}_{p,n-1}^w - \hat{f}_{pn}^w = \dot{f}_p^w \prod_{j=1}^{n-1} (1 - Pr_{pj}^w Pc_{pj}^w) - \dot{f}_p^w \prod_{j=1}^n (1 - Pr_{pj}^w Pc_{pj}^w) = \dot{f}_p^w \prod_{j=1}^{n-1} (1 - Pr_{pj}^w Pc_{pj}^w) Pr_{\bar{p}n}^w Pc_{\bar{p}n}^w$$

Taken together, the expected adaptive flow of the alternative path \bar{p} under n^{th} transfer is expressed as follows:

$$\tilde{f}_{\bar{p}n}^w = \begin{cases} \dot{f}_p^w Pr_{\bar{p}n}^w Pc_{\bar{p}n}^w, & \text{for } n = 1 \\ \dot{f}_p^w \prod_{j=1}^{n-1} (1 - Pr_{pj}^w Pc_{pj}^w) Pr_{\bar{p}n}^w Pc_{\bar{p}n}^w, & \text{for } n > 1 \end{cases}$$

Q.E.D.

The transfer failure rate at j^{th} transfer station on path p before reaching the destination can be given by:

$$Pc_{pj}^w = P_m^{kk'} b_{kp} b_{k'p} \tau_m^{kk'} \psi_{pj}^m \quad (23)$$

where $P_m^{kk'}$ denotes the transfer failure rate from line k to k' at transfer node m . $\psi_{pj}^m = 1$ if node m

is the j^{th} transfer station on path p between OD pair w , and 0 otherwise.

To reflect the rerouting decision, we assume that in case of missed connection, passengers would compare travel cost of the alternative path to that of the currently used path according to their perceptions to decide whether to switch to another alternative, which leads to a switching rate.

The switching rate can be estimated deterministically or stochastically. For the deterministic case, the choice could be modelling based on the travel time difference across alternative routes. In the stochastic version of the alternative route selection problem, travellers may perceive the travel time differently. Assuming that the perception errors are random independent and identically distributed (i.i.d.) variables following a Gumbel distribution, a multinomial logit model can be used to estimate the switching rate under connection failure, which is defined as the ratio between the dis-utilities of the acceptable downstream paths. Note that various assumptions could be made to model the switch rule and the extension would be straightforward.

$$Pr_{pj}^w = \frac{\exp(-\theta U_{pj}^w)}{\exp(-\theta U_{pj}^w) + \exp(-\theta U_{\bar{p}j}^w)} \quad (24)$$

where θ is the degree of passengers' perception of the path travel time.

For sake of parsimony, the travel disutility is measured in generalized time units. Let us denote by R_{pj}^w and $R_{\bar{p}j}^w$ the set of next path elements (links or nodes) starting from j^{th} transfer station on path p between OD pair w . There are two possibilities for the alternative path: (A) the alternative path and originally planned path share the same bus line as illustrated in Fig.1, and passengers only need to stay on board for a holding time at the transfer station; (B) passengers take the other synchronized bus line as an alternative. Note that for a large-scale network a transfer hub may involve with a cluster of attractive routes for transferring passengers to choose from. Under such circumstance, the set of alternative routes can be generated by k-shortest path algorithm with limiting transfers, for instance the total transfer times cannot be more than twice. Therefore, the downstream travel time for designated alternative route is expressed as follows:

$$U_{\bar{p}j}^w = \sum_{k \in R_{\bar{p}j}^w} t_k + \sum_{k \in R_{\bar{p}j}^w} \sum_{k' \in R_{\bar{p}j}^w} (T_c^{kk'} + T_f^{kk'} + Z_m^{kk'}) b_{k\bar{p}} b_{k'\bar{p}} \tau_m^{kk'} \quad (25)$$

On the other hand, passengers who stick to currently used path have to wait for the next bus, thus the downstream travel time also account for the extra delay time. This can be estimated as the departure headway $h_{L_{pj}^w}$, where link L_{pj}^w represents the link immediately connected to the j^{th} transfer station on path p between OD pair w . As a result, the downstream travel time for the originally planned route is obtained, similarly as for Eq.(27), by adding the extra waiting time, i.e.,

$$U_{pj}^w = \sum_{k \in R_{pj}^w} t_k + \sum_{k \in R_{pj}^w} \sum_{k' \in R_{pj}^w} (T_c^{kk'} + T_f^{kk'} + Z_m^{kk'}) b_{kp} b_{k'p} \tau_m^{kk'} + h_{L_{pj}^w} \quad (26)$$

Now we derive the formulations of transfer demand for delay connection taking into account both the non-adaptive and adaptive path flow. Different to the scenario for missed connection, for delay connection scenario, the bus service on connected links are coordinated successfully and passengers would stick to their

current path, such that the non-adaptive flow is identical to the incoming flow before transfer.

By a modification of Eq. (6) to consider the rerouting behaviour, we have the following formulation of transfer demand for delay connection:

$$Q_t^{kk'} = \sum_w \sum_p \sum_n \vec{f}_{pn}^w b_{kp} b_{k'p} \tau_m^{kk'} + \sum_w \sum_{\bar{p}} \sum_n \tilde{f}_{\bar{p}n}^w b_{k\bar{p}} b_{k'\bar{p}} \tau_m^{kk'} \quad (27)$$

where the first term stands for the total incoming flow traversing from line k to line k' , and the second term stands for adaptive flow from other paths that is identical to that of Eq.(22).

Proposition 3. Since passengers travelling along a path may experience a few transfer activities before reaching the current transfer station, the expected incoming flow routing into a specific transfer station should account for the loss flow (by rerouting effect) at preceding transfer stations. As a result, the expected incoming flow entering n^{th} transfer station (of path p between OD pair w) can be computed as follows:

$$\vec{f}_{pn}^w = \begin{cases} \dot{f}_p^w, & \text{for } n = 1 \\ \dot{f}_p^w \prod_{j=1}^{n-1} (1 - Pr_{pj}^w Pc_{pj}^w), & \text{for } n > 1 \end{cases} \quad (28)$$

Proof. When $n = 1$, the incoming flow of path p between OD pair w can be simply represented by the original flow, that is,

$$\vec{f}_{p1}^w = \dot{f}_p^w$$

For path p between OD pair w , the expected flow before undergoing the second transfer is identical to the expected flow after undergoing the first transfer. With the principle of conservation of flows, the expected flow after undergoing the first transfer should be the difference between the expected flow and the rerouting flow.

$$\vec{f}_{p2}^w = \vec{f}_{p1}^w - \vec{f}_{p1}^w Pr_{p1}^w Pc_{p1}^w = \vec{f}_{p1}^w (1 - Pr_{p1}^w Pc_{p1}^w) = \dot{f}_p^w (1 - Pr_{p1}^w Pc_{p1}^w)$$

Similarly, when $n = 3$, the expected flow before undergoing the third transfer is equal to the expected flow after undergoing the second transfer.

$$\vec{f}_{p3}^w = \vec{f}_{p2}^w - \vec{f}_{p2}^w Pr_{p2}^w Pc_{p2}^w = \vec{f}_{p2}^w (1 - Pr_{p2}^w Pc_{p2}^w) = \dot{f}_p^w (1 - Pr_{p1}^w Pc_{p1}^w) (1 - Pr_{p2}^w Pc_{p2}^w)$$

Therefore, the expected incoming flow for n^{th} transfer can be calculated as follows:

$$\vec{f}_{pn}^w = \dot{f}_p^w \prod_{j=1}^{n-1} (1 - Pr_{pj}^w Pc_{pj}^w)$$

Taken together, the expected incoming flow of path p between OD pair w under n^{th} transfer can be expressed by the following piecewise function:

$$\vec{f}_{pn}^w = \begin{cases} \dot{f}_p^w, & \text{for } n = 1 \\ \dot{f}_p^w \prod_{j=1}^{n-1} (1 - Pr_{pj}^w Pc_{pj}^w), & \text{for } n > 1 \end{cases}$$

Q.E.D.

Combining Eq.(30) with Eq.(23), we have:

$$\vec{f}_{pn}^w = \frac{\tilde{f}_{pn}^w}{Pr_{pn}^w Pc_{pn}^w} \quad (29)$$

Similarly, with rerouting behaviour, the through demand with rerouting behaviour takes the following form:

$$Q_{mk} = \sum_w \sum_p \sum_n \tilde{f}_{pn}^w \varphi_k^m b_{kp} + \sum_w \sum_{\bar{p}} \sum_n \tilde{f}_{\bar{p}n}^w \varphi_k^m b_{k\bar{p}} \quad (30)$$

In Eq.(32), the first term stands for the through demand on the original path, and the second term stands for the rerouting demand swapping from other paths.

3.4.4 Extension to multiple transfer opportunities

In a transit network, during a trip, a rider could have multiple transfer opportunities and they may modify their choice of transfer point (thus path choice) multiple times (as shown in Fig.7). To capture this effect, the adaptive flow of the alternative path \bar{p} under n^{th} transfer, $\tilde{f}_{\bar{p}n}^w$, should be redistributed. In this sense, the adaptive flow $\tilde{f}_{\bar{p}n}^w$ can be treated as “virtual” domain path flow, and is reassigned once the passengers undergoing a transfer point.

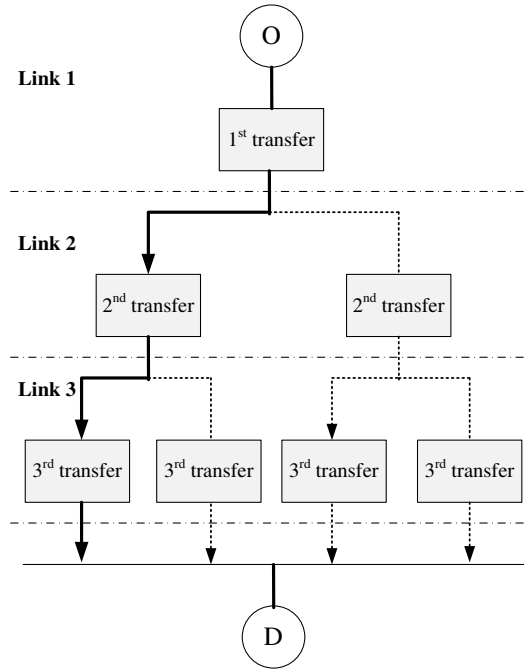


Fig.7 Illustration of stochastic rerouting process with multiple transfer opportunities

Now we discuss the derivation for $\tilde{f}_{\bar{p}n}^w$. To do so, we introduce a set of variables $\hat{f}_{\bar{p}nr}^w$ and $\tilde{f}_{\bar{p}nr}^w$ to represent the r^{th} transfer point on the alternative path \bar{p} . Then the formulations could be obtained following proposition 1.

$$\hat{f}_{\bar{p}nr}^w = \tilde{f}_{\bar{p}n}^w \prod_{r=1}^n (1 - Pr_{\bar{p}r}^w Pc_{\bar{p}r}^w)$$

$$\tilde{f}_{\bar{p}nr}^w = \begin{cases} \tilde{f}_{\bar{p}n}^w Pr_{\bar{p}r}^w Pc_{\bar{p}r}^w, & \text{for } r = 1 \\ \tilde{f}_{\bar{p}n}^w \prod_{j=1}^{r-1} (1 - Pr_{\bar{p}j}^w Pc_{\bar{p}j}^w) Pr_{\bar{p}r}^w Pc_{\bar{p}r}^w, & \text{for } r > 1 \end{cases}$$

where the transfer failure rate Pc_{pr}^w and switching rate Pr_{pr}^w can be obtained following Eqs.(23) and (24).

As we can see, the extension is straightforward and thus the model is scalable. However, in a real situation, passengers would generally reach their destinations with no than three times of transfers (Yu et al, 2015; Xu et al, 2017), therefore it is reasonable to assume that passengers would not modify their transfers for more than twice.

3.5. Bi-level formulation of the schedule coordination problem

The objective of the proposed model is two-fold: one aims to minimize the systematic total costs, including the user cost and operation cost, while the other aims to capture the passengers' pre-trip and on-trip route choice behaviour. This gives rise to a master-slave problem: passenger path choice decisions are influenced by schedule coordination schemes, and schedule coordination is dependent on passenger demands. Therefore, the problem can be formulated as a bi-level programming model. The objective function for the upper level optimization is to find out the bus schedule coordination scheme minimizing the total costs. The optimized coordination solution is passed to the lower level model. Based on a given schedule coordination scheme, the lower level model calculates the expected flow patterns. The system converges when it minimizes the total costs.

The framework of the proposed bi-level programming is presented in Fig. 8.

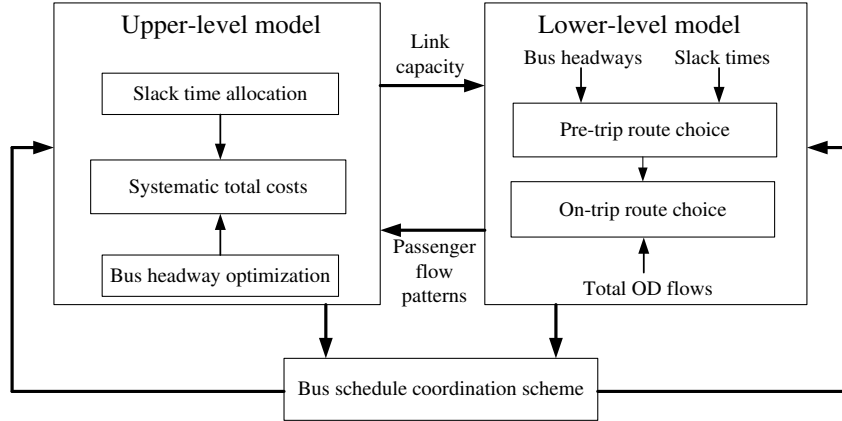


Fig.8 The framework of the bi-level program

3.5.1 Upper-level schedule coordination problem

The upper-problem is to determine the headways h_k and slack times s_{mk} that minimize the total costs. In this study, headways are optimized as integer multiples of the base cycle. Let the base cycle y belong to a set consisting of several discrete values, i.e., $y \in Y$, $Y = \{y_1, \dots, y_i, \dots, y_I\}$. Similar to Kim and Schonfeld (2014), for realism we assume that y is selected among $\{2, 3, 4, 5, 6, 7.5, 10, 12, 15, 20 \text{ and } 30 \text{ min}\}$. Let δ_i denote an indicator representing whether y_i is selected, then the optimization problem can be converted from exploring optimal combination of (y, β_k, s_{mk}) to optimal combination of (b_i, β_k, s_{mk}) . Thus, the upper-level problem is formulated into a 0-1 mixed-integer nonlinear programming model as follows:

$$\min C(b_i, \beta_k, s_{mk}) \quad (31)$$

subject to

$$h_k = \beta_k \sum_{i=1}^I y_i b_i \quad (32)$$

$$h_k^{min} \leq h_k \leq h_k^{max}, \forall k \quad (33)$$

$$h_k \in Z^+ \quad (34)$$

$$y_i \in Y \quad (35)$$

$$b_i \in \{0,1\} \quad (36)$$

$$\sum_{i=1}^I b_i = 1 \quad (37)$$

$$\beta_k \in Z^+, \forall k \quad (38)$$

$$s_{mk} \geq 0, \forall k, m \quad (39)$$

The objection function (31) minimizes the system total costs. Eq.(32) guarantees that the line-specific headways are integer multiples of the base cycle. Eq.(33) states that the headway should range between a minimum and a maximum acceptable headway. Eq.(34) guarantees that headways are integer values. Eq.(35) represents that the base cycle is chosen from a set of discrete values. Eq.(36) defines b_i as binary variables, if y_i is selected, then $b_i = 1$, otherwise, $b_i = 0$. Eq.(37) guarantees that only one element in the set is selected. Eq.(38) guarantees that the ratios are positive integer values. Eq.(39) guarantees that the slack times should be positive values.

3.5.2 Lower-level problem

The lower level model is used to describe the passengers' route choice behaviour under schedule coordination environment. In the first stage, all users in the system make pre-trip decisions in an attempt to minimizing their travel costs. This can be also approximated as the equilibrium state. In the second stage, the user is confronted with a missed connection upon transfer and he is left with a decision whether to wait until the next bus arrives or to take a detour route as an alternative. As a result, the flow patterns should be redistributed to account for the connection uncertainty and rerouting effect (As presented in section 3.4).

Now we discuss the calculation of flow distribution in the first stage, more specifically the original path flow f_p^w . Since bus line headways are decision variables in the optimization model, it is natural to model the flow distribution of pre-trip route choice as a frequency-based transit assignment. It is assumed that each passenger would make the trip plan according to the expectation for travel costs. We use a utility-based approach to describe traveller's perceived travel costs.

As a result, the expected travel disutility U_p^w for any path p between OD pair w in the transfer network can be calculated as the sum of all of expected travel disutilities on links along this path, including the initial waiting time, in-vehicle time, and expected transfer waiting time.

$$U_p^w = EW_s \xi_{sp}^w + \sum_s T_s \delta_{sp}^w + \sum_m \tilde{T}_m^p \quad (40)$$

where \tilde{T}_m^p represents the transfer waiting time cost at transfer node m on path p .

Most of the existing transit assignment models without schedule coordination that assume an arbitrary constant (usually a large number) as a transfer penalty (e.g., Yao et al, 2012; Szeto et al, 2013). In the perfect (deterministic) schedule coordination, passengers could transfer smoothly between different bus lines.

However, in the context of stochastic schedule coordination with uncertain disruptions, the transfer penalty would be dependent on the travel time variability and the schedule coordination scheme (headways and slack times), as well as the resulting expected transfer waiting time. Since there may exist multiple combinations of connecting lines at a transfer station, the expected transfer waiting time at transfer node m could be approximated by the weighted average of transfer waiting time of connecting bus lines along the path. The expected transfer waiting time includes missed connection time, delayed connection time and (if any) the inter-cycle transfer waiting time:

$$\tilde{T}_m^p = \sum_k \sum_{k'} \pi_s^k (T_c^{kk'} + T_f^{kk'} + Z_m^{kk'}) b_{kp} b_{k'p} \tau_m^{kk'} \quad (41)$$

As a result, the network flow patterns before rerouting under given schedule plan could be obtained by the transit assignment models. With the original (before rerouting) flow patterns, the expected flow distribution with rerouting behaviour can be further obtained by Section 3.4.2 to adapt to changing operational condition.

3.5.3 Solution algorithm

Given a set of headways and slack times, users will choose acceptable paths prior to starting their journey and reroute in case of missed connection. Likewise, given a set of flow patterns selected by system users, system designers can re-allocate the vehicle resources to avoid extra costs caused by congestion. Then, users will choose the new trip chains again. Therefore, we can use this iterative strategy between the upper-level and lower-level decisions to reach the optimal solution gradually. The method of successive averages (MSA) is used to address the demand assignment problem under a given schedule design, (Poon et al., 2004; Si et al., 2016). More specifically, the pseudo code of heuristic algorithm is presented as follows:

Step 1: Initialization. Let iteration index $z = 0$ and the initial slack times be 0 (i.e., $s_{mk}^z = 0$), and initialize headways h_k^z as a combination of positive values subject to Eq. (32) to (39).

Step 2: Update. Apply the solution of slack times and headways into the lower-level model and solve it using MSA. Obtain a set of solution of passenger path flows $f_p^{w,z+1}$ and the resulting link flow x_k^{z+1} . Subsequently, the expected network flow after redistribution can be obtained.

Step 3: Modification. Apply the value of passenger flows $f_p^{w,z+1}$ solved in Step 2 into the upper-level model. Then, search a set of optimum solution of slack times and headways s_{mk}^{z+1} , h_k^{z+1} .

Step 4: Convergence criteria. Check whether the convergence criterion is satisfied or not. The algorithm terminates if the relative gap of traffic flow between the iterations becomes less than a threshold; otherwise, let $z = z + 1$, and continue the iteration between Step 2 and 3.

4. Numerical example

To evaluate the influence of rerouting to schedule planners, a small numerical test is firstly conducted to compare the performances of enhanced schedule coordination with rerouting with the base case without rerouting. Then, the influence of different operating settings on the model performance is discussed. Lastly, we apply the model to a medium-size hybrid bus network and analyse the passengers' rerouting behaviour

on schedule coordination.

4.1. Experiment I

We consider the example corresponding to the network illustrated in Section 3.1 reproduced in Fig.9, the hypothetical triangular loop bus network is a three-route network containing 18 links and 9 nodes, of which 3 nodes are transfer stations. For such a network, all lines interact with each other so that the majority of passengers have more than one route option to get to their destinations. The extension of this network to a general bus network with more immediate stops would be straightforward. The OD demand matrix for the network is asymmetrical and presented in Table 2, and the information of link travel times are summarized in the brackets in Fig.7 (the bidirectional link travel times are assumed to be identical).

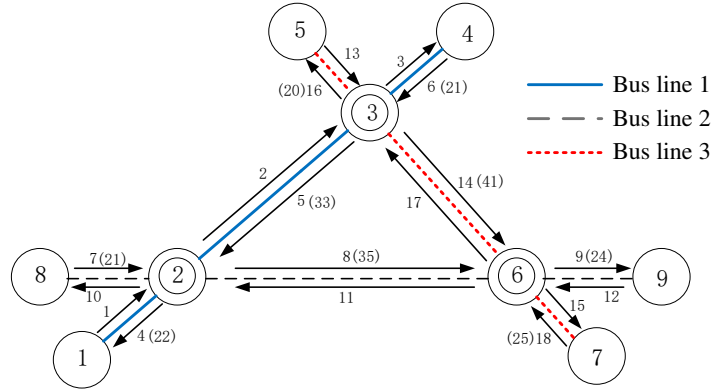


Fig. 9 A triangular loop network

The transfer reliability mainly depends on bus travel time variability and slack times. Following the literature, the exponential distribution is employed to model the arrival delay of buses (Wu et al, 2016; Randolph, 2001; Bookbinder and Desilets, 1992). The probability density function and the respective cumulative density function of arrival delay are expressed as follows, which is used in the calculation of transfer waiting time costs:

$$f(d_{mk}) = \begin{cases} \lambda_{mk} e^{-\lambda_{mk} d_{mk}}, & d_{mk} > 0 \\ 0, & d_{mk} \leq 0 \end{cases} \quad \text{and} \quad (42)$$

$$F(d_{mk}) = \begin{cases} 1 - e^{-\lambda_{mk} d_{mk}}, & d_{mk} > 0 \\ 0, & d_{mk} \leq 0 \end{cases}$$

where d_{mk} is the arrival delay of buses; and parameter $\lambda_{mk} > 0$ is the parameter of the exponential distribution.

For exponential distribution, the mean delay and the variation of delay are $1/\lambda_{mk}$ and $1/\lambda_{mk}^2$, respectively. And the delay probability decreases quickly as the delay time increases. With the exponential delay arrival assumption, the computation time can be effectively reduced by analytical formulations featuring integrable functions. However, in principle, the model can be applied to any distribution of delayed arrive time.

The unit operation cost in Eqs.(1) and (19) is taken as $B = 0.667\$/veh/min$. The vehicle capacity Ca is taken as 35 pax. Two values of waiting time are taken as $\mu_c = \mu_w = 0.2\$/min$. The delayed parameter

in Eq.(42) is set to be $\lambda_{mk} = 0.25min^{-1}$. The other default model parameters used are: $\theta = 0.9$, $a = 1$, and $b = 4$.

Table 2. The O/D demand matrix for the sample network (pax/min)

Origin	Destination								
	1	2	3	4	5	6	7	8	9
1	0.00	0.1	0.06	0.06	0.1	0.04	0.08	0.06	0.04
2	0.10	0.00	0.08	0.08	0.00	3.00	0.06	3.00	0.00
3	0.06	0.05	0.00	0.04	0.04	0.06	0.03	0.03	0.05
4	0.06	0.04	0.04	0.00	0.04	0.02	0.02	0.00	0.02
5	0.10	0.08	0.08	0.08	0.00	0.06	0.06	0.06	0.00
6	0.04	2.00	0.06	0.02	0.02	0.00	0.00	1.00	2.00
7	0.08	0.1	0.06	0.06	0.06	0.00	0.00	0.08	0.06
8	0.03	2.5	0.03	0.00	0.01	0.50	0.03	0.00	0.50
9	0.03	1.5	0.05	0.01	0.00	1.50	0.50	0.50	0.00

Solving the optimization problems for the baseline values yields the results shown in Table 3. The outcomes include the optimal headways and slack times, and the individual cost components as well as the resultant total costs, distinguished by the base and enhanced version. The base and enhanced version represents the schedule coordination model without and with rerouting, respectively. One can see that, compared to the base model, it is more beneficial to operate with longer headways for the enhanced model, leading to overall lower operation cost C_o and induced slack time cost C_s . Nevertheless, the transfer costs (C_c , C_f , C_g) for the enhanced model have been reduced in comparison to the base model. This is because, as discussed previously, the number of passengers suffering from connection failure could be reduced when rerouting is allowed. Although the incorporation of rerouting results in a longer headway (and therefore longer transfer waiting time), this has been overwhelmed by the reduced number of transfer passengers who have to wait for a full headway. As a result, the total transfer waiting time, which equals to the transfer waiting time multiplied by the transfer demand, could still be reduced to some extent. This indicates that incorporation of rerouting behaviour leads to longer headways (thus yields fleet size saving) without increasing passengers' costs.

Table 3. Optimized results for the baseline values

Variables (min)	Base	Enhanced	Cost	Base	Enhanced	
			(\$/min)			
$h_k (h_1, h_2, h_3)^a$	(24, 12, 24)	(28, 12, 28)	C	146.53	143.74	
s_{21}	(0.88, 1.93)	(0.77, 1.84)	C_o	16.40	15.11	
s_{22}	(0.25, 0.00)	(0.01, 0.02)	C_w	26.04	26.60	
$s_{mk}(\text{note})$	s_{31}	(3.63, 1.62)	(3.91, 1.61)	C_f	91.90	91.92
	s_{33}	(2.29, 3.76)	(2.73, 4.18)	C_s	1.17	0.92
	s_{61}	(0.38, 0.40)	(0.10, 0.03)	C_c	4.31	2.61

s_{63}	(1.68, 1.31)	(1.60, 1.07)	C_f	4.57	4.34
	—		C_g	2.14	2.24

Note: The two values in the bracket denote the slack times for the two different directions of the line.

Next, we conduct sensitivity analysis to investigate the relationship between model parameters and decision variables, and discuss the implications for transit operation and network design. Since the arrival delay at transfer stations are central sources triggering connection failure and adaptive route choice, in what follows we begin with investigating the sensitivity to delay arrival uncertainty, followed by the sensitivity to total demand. It proceeds to analyse the effect of rerouting on the system performance under supply constraints, which is common practice in transit operation. To achieve this, slack times are optimized for exogenously specified headways.

4.1.1 Sensitivity to arrival mean delay

Fig. 10 presents the averaged optimal slack time under various levels of arrival delay. It can be seen that the averaged slack times for both models tend to increase initially with the mean delay before decreasing. This is because that the delay arrival is assumed to be exponentially distributed, the variance of delay $1/\lambda^2$ also increases with the mean delay $1/\lambda$. This suggests that the system needs more slack time when the uncertainty is higher. However, adding more slack times would be undesirable when the arrival delay uncertainty is higher than a threshold.

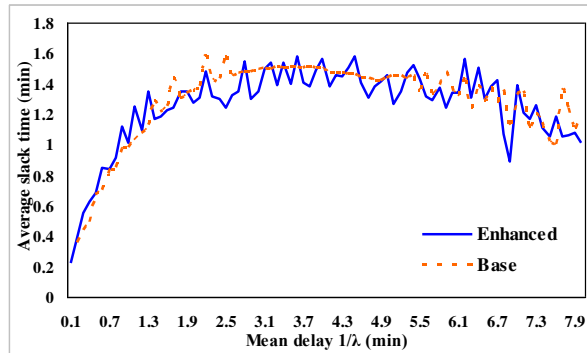


Fig. 10 Averaged optimal slack time under different mean delays

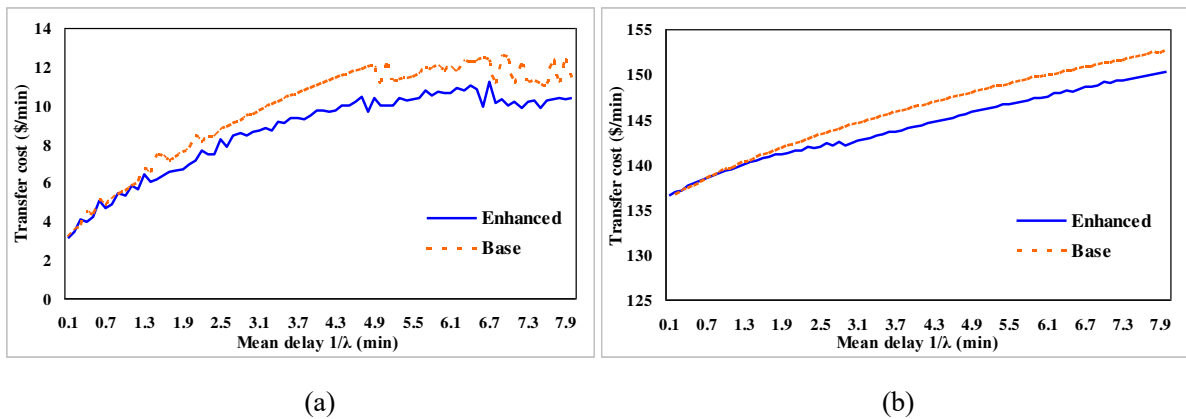


Fig. 11 System costs under various mean delays: (a) transfer cost; (b) total cost

Fig. 11 shows the key cost components of schedule coordination under different mean delays: total transfer cost ($C_c + C_f + C_g$) and total system cost (C). It is found that the transfer cost and systematic total cost generally increase as the mean delay increases. However, there is a sudden drop in transfer cost when the mean delay is higher than 4.3 min. This is because the increase of the mean delay changes the optimized headways, which in turn affects the transfer cost. The costs of the enhanced model are consistently lower than those of the base one, and the gap between them is negligible when the mean delay $1/\lambda$ is lower than 1.3. The reason is that the probability of missed connection is higher when the arrival delay becomes larger, under which circumstance the rerouting behaviour takes more effect. Accordingly, lower system total cost can be expected.

4.1.2 Sensitivity to total demand

Since bus departure headways rest with the demand for the service, here we first analyse the optimized headways of both base and enhanced models at different demand levels, and the results are presented in Table 4. The first observation is that schedule coordination with integer-ratio headways outperforms that with a single common headway in all scenarios. This implies that the demand or route length is significantly different among the bus routes. As the demand level grows proportionally, the optimal headways decrease, and the optimal headways of Lines 1 and 3 versus Line 2 remain no less than 2 before the demand ratio reaches a critical value (1 and 1.4 for the base and enhanced case, respectively). Beyond that, the ratio between them becomes smaller than 2. This indicates that the integer-ratio headway may be more economical for schedule coordination at low demand. In addition, we observe that the headways are reduced after incorporating rerouting behaviour in the model, suggesting that the fleet size can be saved and thus economically advantageous. This is because allowing rerouting will mitigate the negative effect of missed connection, less buses then are needed.

Table 4. Optimized headways at different demand ratios (min)

Ratio	Base			Enhanced		
	h_1	h_2	h_3	h_1	h_2	h_3
0.2	56	28	56	60	30	60
0.4	40	20	40	45	20	45
0.6	32	16	32	35	15	35
0.8	30	15	30	32	16	32
1	24	12	24	28	12	28
1.2	21	12	21	27	12	27
1.4	21	12	21	21	12	21
1.6	18	12	18	20	12	20
1.8	16	12	16	18	12	18
2.0	16	12	16	18	12	18

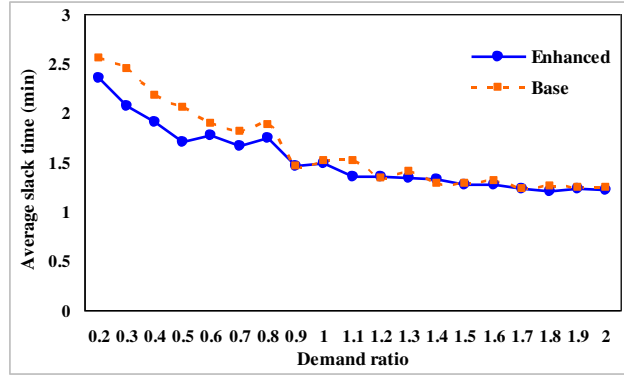


Fig. 12 Averaged optimal slack time under different demand levels

Fig.12 presents the averaged optimal slack time under different demand levels. We observe that the slack times decrease with the increment of the demand volume in a nearly exponential way. This is expected since at higher demand, the service headway will be shorter and any missed connection becomes less frustrated for transfer passengers, thus less slack times are needed to relief the travel time variability. In addition, at low demand, the slack times for the enhanced model are smaller than that of the base one, and the gaps are negligible when the demand ratio is larger than a critical value (e.g., 1.2 in this example). There are two possible reasons for this. First, the slack times are linked closely to service headways, while the headways for both variants are close at relatively heavy demand as shown in Table 6. Second, according to Eq.(28), as the headway becomes shorter, the travel time of the originally planned route would be reduced in spite of connection failure, such that passengers are less likely to modify their routes (and therefore the switching rate has been reduced).

4.1.3 Sensitivity analysis under supply constraints

While it is shown that incorporating rerouting in the optimization of schedule coordination without any supply constraints could lead to cost saving, it is not uncommon for a transit operator to face with supply constraints in practice due to budget or fleet size limitation. Therefore, from the standpoint of operator, it is important to understand the schedule performance with such consideration. To further investigate whether the enhanced model is advantageous under supply constraints, slack times are optimized for exogenously specified headways. This resembles finding the optimal schedule plan in the context of a constant fleet size. For simplicity and clarity, we focus on verifying the slack times and system costs under the optimal headway for base demand, that is, fixing headways at $h_1 = 24min$, $h_2 = 12min$, $h_3 = 24min$.

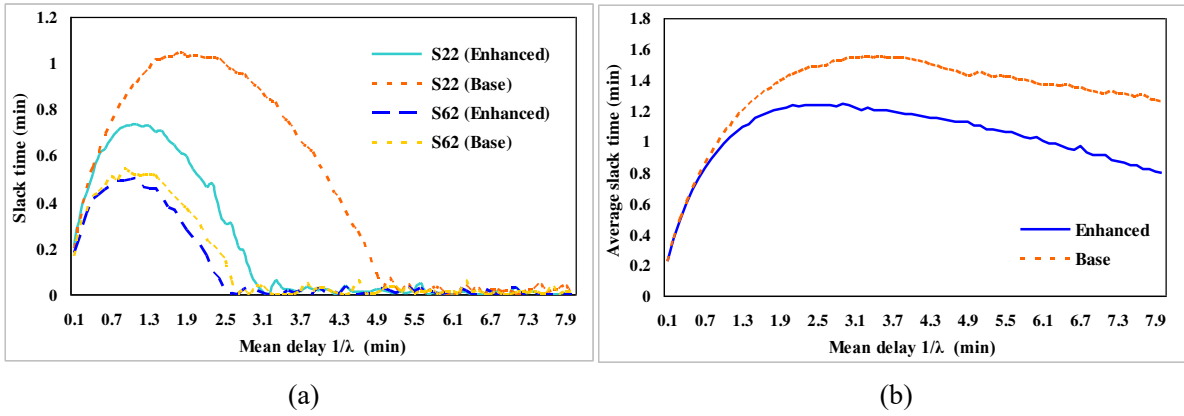


Fig.13 Optimal slack times for various mean delays: (a) typical slack time; (b) mean slack times

Fig.13 presents the typical optimized slack times (s_{22} , s_{62}) and averaged slack time for various mean delay. It is found that the mean slack times are smoother than in Fig.10, which is because the slack times are optimized with predetermined headways. As the mean delay increases, the typical slack time shows an upward trend before hitting the threshold and then decreases drastically to 0. More importantly, the typical slack times and average slack time characterized by rerouting effect are considerably shorter than those of the base model. The saving for the average slack time made by rerouting appears to get enlarged when the mean delay is larger than 1.1. This may be explained by the fact that the slack time setting is related to the trade-off between the transfer demand and through demand, while as per discussed, the immediate effect of passenger rerouting is the reduction of transfer demand and the increase of through demand. Since rerouting is triggered by transfer failure, more chance of connection failure and rerouting can be expected when the arrival delay uncertainty is higher. As mentioned before, the reservation of large slack times is inefficient since it leads to additional vehicle holding and trip time. In other words, the allocated slack time on the other hand means the efficiency of schedule coordination. Therefore, these results suggest that by introducing the rerouting behaviour, more cost-effective synchronized schemes with less slack times can be realized.

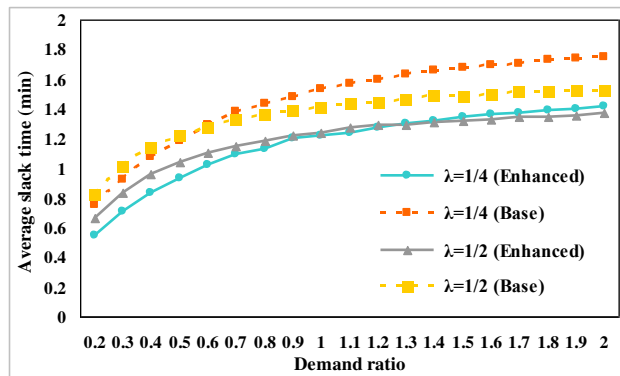


Fig. 14 Averaged optimal slack time for different demand levels under given headways

We further analyse the change pattern of average slack times for different demand ratios under given headways and the results are shown in Fig.14. The scenarios differ in two dimensions: whether the rerouting is allowed and arrival delay uncertainty is high or moderate. It is clear that the average slack times increase

with the demand level. This is because transfer volume grows proportionally with the total demand, and when headways are fixed, more slacks are needed to relieve the travel time variability and reduce the transfer costs. Interestingly, the rerouting effect could reduce slack time reservation under various demand ratio, and the improvement is greater for the high delay scenario ($\lambda = 1/4$). For example, the slack time saving for high delay scenario are 37.3% and 23.1% for 0.2 and 2 demand ratio, respectively; whereas for moderate delay scenario are 24% and 11%, respectively. (In this study the saving is calculated as $(nonadaptive - adaptive) \times 100 / adaptive$). This is explainable from the fact that rerouting effect is more active under higher delay arrival uncertainty as discussed previously. These results give us a practical managerial insight that, in the context of supply constraints, rerouting can improve performance by a greater degree when the arrival delay is higher and demand level is lower.

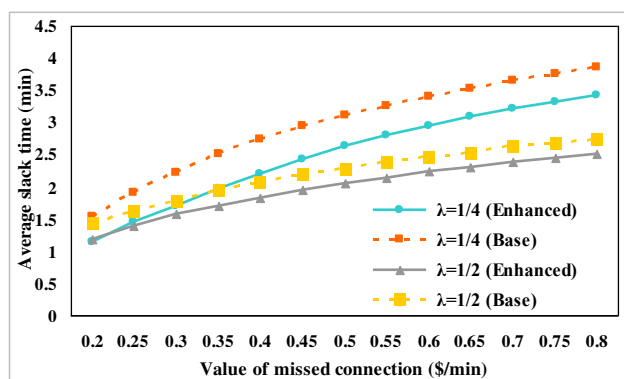


Fig.15 Averaged optimal slack time versus value of missed connection time

Similarly, by increasing the value of missed connection time (μ_c) while fixing the values of other user's time components, the average optimal slack times are sketched in Fig.15. One can see that the average slack times increase as the value of missed connection time increases, suggesting that coordinated transfer should be encouraged. Similar as shown before, the slack times of enhanced model are considerably smaller than those of the base one, and the gaps between them are larger when the arrival delay uncertainty is higher.

4.2. Experiment II

This section analyses the passenger flow pattern with rerouting behaviour. Consider a medium-size hybrid network shown in Fig.16 with 4 bus lines and 3 transfer stations. Stations 1 to 3 are common line stops shared by line 1 and line 2. The transfer coordination at the immediate stations along the common line are ignored since in reality the link travel time might be different across bus lines and it required additional effort in synchronization timetable from the terminal as suggested by Ibarra-Rojas et al (2012) and Ibarra-Rojas and Muñoz (2016), which is out of the scope of this study. This network can be viewed as an extended version of the simple triangular network as shown in Fig.6, where line 1, 2 and 3 constitutes a triangular loop network while line 3 combined with line 4 are equivalent to a trunk-and-feeder network. The trip rate for each OD pair is assumed to be 0.5pax/min, and the link travel times are provided in Table 5. For simplicity the link travel times in both directions are assumed to be identical. Once again the exponential distribution is used for delayed arrival distribution function, of which in the base case the delay parameter $\lambda_{mk} =$

$1/8min^{-1}$ is fixed for all transfer nodes and on all lines.

Table 5. Link travel times (min) for the bus network

Link	1-2	2-3	3-4	4-5	5-6	6-7	4-8	8-9	8-10	10-11
Travel time	5	6	5	4	6	5	7	6	5	4
Link	5-8	5-12	12-13	13-14	13-18	16-17	17-18	13-19	19-20	20-21
Travel time	6	7	4	5	7	5	8	6	6	5

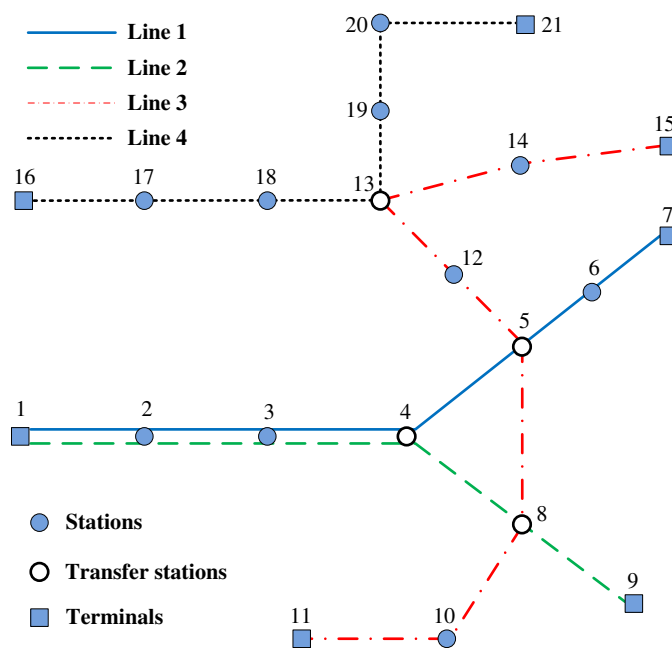


Fig.16 A medium-size hybrid bus network (adapted from Xu et al., 2017)

Table 6. Passenger flow distribution for some typical O-D pairs

O-D pair	Path node sequence	Path flow	Rerouting flow	Rerouting flow/Path flow
15→1	15,14,13,12,5,4,3,2,1	0.348	0.007	2.03%
	15,14,13,12,5,8,4,3,2,1	0.152	-	-
15→4	15,14,13,12,5,4	0.346	0.041	11.8%
	15,14,13,12,5,8,4	0.154	-	-
16→2	16,17,18,13,12,5,4,	0.347	0.033	9.6%
	16,17,18,13,5,8,4	0.153	-	-
16→3	16,17,18,13,12,5,4,3	0.345	0.041	11.9%
	16,17,18,13,12,5,8,4,3	0.155	-	-
17→2	17,18,13,12,5,4,3,2	0.346	0.033	9.67%
	17,18,13,12,5,8,4,3,2	0.154	-	-
17→4	17,18,13,12,5,4	0.344	0.041	11.9%
	17,18,13,12,5,8,4	0.156	-	-

To illustrate how the passenger flow distribution changed by rerouting behaviour, the detailed results for

some typical O-D pairs are analysed. There are two paths for these O-D pairs, and their alternative paths are branched from a transfer station (node 5). For example, one of the paths for OD pair 15→1 is line 3→line 1 transferring at node 5, while the alternative path is line 3→line 2 transferring at node 8. Table 6 presents the originally planned path flow and the rerouting flow for typical O-D pairs. One can see that a considerable proportion of the passenger flow is migrated into the alternative path with connection uncertainty, ranging from about 2% to as high as 12%.

Table 7. Rerouting flow factor after doubling mean arrival delay at node 5

O-D pair	Factor $\hat{f}(\lambda^{-1})/\hat{f}(2\lambda^{-1})$	$\hat{f}(2)/\hat{f}(4)$	$\hat{f}(4)/\hat{f}(8)$	$\hat{f}(8)/\hat{f}(16)$
15→1		0.831	0.957	0.987
15→4		0.802	0.951	1.026
16→2		0.801	0.946	1.030
16→3		0.803	0.952	1.026
17→2		0.802	0.947	1.031
17→4		0.804	0.953	1.030

As missed connection is the main source triggering rerouting behaviour, how the change of vehicle arrival delay parameter affects the passenger flow distribution deserves some discussion. To this end, we investigate the evolution of rerouting flow factor, i.e., $\hat{f}(\lambda^{-1})/\hat{f}(2\lambda^{-1})$, when the delay parameter $1/\lambda$ at node 5 is doubled and the results are shown in Table 7. One can see that, as the delay parameter increases, the factor tends to increase, indicating that more passenger flow is migrated into the alternative path when the delay parameter becomes higher. When the delay parameter reaches a threshold, i.e., $1/\lambda = 8$, the factor $\hat{f}(8)/\hat{f}(16)$ is very close to 1. As mentioned before, since the delay arrival time is assumed to follow exponential distribution, its variance $1/\lambda^2$ will increase as the mean delay $1/\lambda$ increases. This implies that 8 is already a high delay parameter, and any increment of rerouting flow could become more difficult though increasing the parameter.

With transfer failure and the rerouting flow, demand transition may occur from transfer passengers to through passengers at a designated transfer station, which further results in the change of the schedule coordination planning. This supports the conjecture mentioned in Fig.5 that rerouting could lead to the change in transfer demand (and thus through demand). Therefore, the results suggest that passenger rerouting cannot be ignored in the context of schedule coordination and connection uncertainty.

5. Concluding remarks

Rerouting can efficiently deal with uncertainties. However, the issue of bus schedule coordination design with rerouting of passengers has not been fully addressed before. This study proposed a new model for stochastic bus schedule coordination design, providing that passengers can adjust their route in case of missed connection. To build the linkage between the expected network flows and transfer reliability, we developed a bi-level programming model wherein the schedule design (headways and slack times) and

passenger route choice are determined simultaneously. The upper-level problem is designed to minimize the system total costs (user cost and operator cost), while the lower-level problem is a route choice (pre-trip and on-trip) model for timed-transfer service. A heuristic algorithm and MSA have been jointly applied for solving the bi-level model.

This model was exemplified by a simple triangular loop bus network and a medium-size hybrid network with ubiquitous rerouting behaviour. Experiments revealed that effect of rerouting should not be ignored in the design of schedule coordination. Results showed that the incorporation of dynamic rerouting behaviour leads to longer headways (thus yields fleet size saving and economic benefit) without increasing passengers' costs. Another interesting finding is that in the context of supply constraints (i.e., given headways), more cost-effective schedule coordination schemes with less slack times can be achieved, and the improvement is greater when the arrival delay uncertainty is higher. These findings indicate that ignoring such behavioral characteristic would induce idle capacity in the transit system, including higher costs for the operators, thus underestimate the efficacy of schedule coordination scheme and lead to potential planning errors. Therefore, the resultant model provides more options (i.e., rerouting) and flexibility for planners to tackle uncertainties more effectively. In view of these findings, we recommend that a properly designed transfer-based network with well-connected routes and real-time information can improve the efficiency of transit system and deal with disruptions more cost-effectively.

This study stepped out an innovative analytical framework for incorporating demand assignment and rerouting behaviour in stochastic schedule coordination design, which paves the way for new research directions. For instance, while this article focuses on the schedule design over a specific planning horizon in which passenger demand and travel time retain their characteristics, future research can extend the lower level model to other more realistic, dynamic transit assignment models, and incorporate demand and running (and delayed) time dynamics in multi-period schedule coordination planning to capture the time-dependent operating conditions. Additionally, the proposed framework could be extended to joint optimization of optimum transit network layout and schedule coordination.

Acknowledgements

This work is partly supported by the National Science Foundation of China (Project No. 61703165, 61473122), the Fundamental Research Funds for the Central Universities (Project No. D2171990), and the Royal Academy of Engineering (project UK-CIAPP\286).

Appendix 1. Derivation of missed connection time $T_c^{kk'}$

The event of missed connection occurs when the connecting bus leaves before the transfer passengers arrive. Assuming the delayed arrival times on different lines are statistically independent, then the joint probabilities of simultaneous arrivals could be attained by multiplying the arrival probabilities of each individual bus line. Let k' denote the bus line connecting to line k at transfer node m (which also holds in the following appendices). There are two possible cases formulated as follows:

Case A. The bus on one line arrives at transfer node behind schedule while the connecting bus on another

line arrives before schedule. The expected missed connection time is:

$$\int_{s_{mk}}^{h_k} \int_0^{s_{mk'}} [h_{k'} - (d_{mk} - s_{mk})] f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk} dd_{mk'}$$

Case B. Both buses arrive at transfer node behind schedule, while the bus on one line arrives after the connecting bus on another line. The expected missed connection time is:

$$\int_{s_{mk}}^{h_k} \int_{s_{ml}}^{d_{mk} - s_{mk} + s_{mk'}} [h_{k'} - (d_{mk} - s_{mk})] f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk} dd_{mk'}$$

Then the expected missed connection time is the sum of the waiting time for these two scenarios:

$$T_c^{kk'} = \int_{s_{mk}}^{h_k} \int_0^{s_{mk'}} [h_{k'} - (d_{mk} - s_{mk})] f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk} dd_{mk'} +$$

$$\int_{s_{mk}}^{h_k} \int_{s_{ml}}^{d_{mk} - s_{mk} + s_{mk'}} [h_{k'} - (d_{mk} - s_{mk})] f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk} dd_{mk'}$$

Appendix 2. Derivation of delayed connection time $T_f^{kk'}$

The calculation of delayed connection cost should account for three possible options:

Case A. Both buses are not late. Then the waiting time is $s_{mk} - d_{mk}$. The expected waiting time could be calculated using integration:

$$\int_0^{s_{mk'}} \int_0^{s_{mk}} (s_{mk} - d_{mk}) f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk'}$$

Case B. The bus on one line arrives before schedule but the bus on another line arrives behind schedule. Then the transfer waiting time for passengers is $d_{mk'} - s_{mk'} + s_{mk} - d_{mk}$. The expected waiting time could be calculated as:

$$\int_0^{s_{mk}} \int_{s_{mk'}}^{h_{k'}} (d_{mk'} - d_{mk} - s_{mk'} + s_{mk}) f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk'}$$

Case C. Both buses arrive behind schedule, but the bus on one line arrives before its connecting bus on another line. Then the transfer waiting time for passengers is $d_{mk'} - s_{mk'} + s_{mk} - d_{mk}$. The expected waiting time is:

$$\int_{s_{mk'}}^{h_{k'}} \int_{s_{mk}}^{s_{mk} - s_{mk'} + d_{mk'}} (d_{mk'} - d_{mk} - s_{mk} + s_{mk'}) f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk'}$$

Consequently, summing up the expected waiting time for all possible options yields the total delayed connection time:

$$T_f^{kk'} = \int_0^{s_{mk'}} \int_0^{s_{mk}} (s_{mk} - d_{mk}) f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk'}$$

$$+ \int_0^{s_{mk}} \int_{s_{mk'}}^{h_{k'}} (d_{mk'} - d_{mk} - s_{mk'} + s_{mk}) f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk'}$$

$$+ \int_{s_{mk'}}^{h_{k'}} \int_{s_{mk}}^{s_{mk} - s_{mk'} + d_{mk'}} (d_{mk'} - d_{mk} - s_{mk} + s_{mk'}) f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk'}$$

Appendix 3. Derivation of transfer failure rate Eq.(21)

The transfer failure rate, or missed connection probability depends on the delayed arrival distribution, and how service headways and slack times are configured. There are two possible outcomes:

Case A. The bus on one line arrives behind schedule at transfer station but the connecting bus on another line arrives before schedule, that is, $d_{mk} > s_{mk}$ and $d_{mk'} < s_{mk'}$. Then the respective probability is:

$$\begin{aligned} & P(d_{mk} > s_{mk} \cap d_{mk'} < s_{mk'}) \\ &= P(d_{mk} > s_{mk})P(d_{mk'} < s_{mk'}) \\ &= \int_{s_{mk}}^{h_k} \int_0^{s_{mk'}} f(d_{mk}) dd_{mk} f(d_{mk'}) dd_{mk'} \end{aligned}$$

Case B. Both buses arrives behind schedule, but the bus on one line arrive before the connecting bus on another line. Then the respective probability is:

$$\begin{aligned} & P(d_{mk} > s_{mk} \cap d_{mk'} > s_{mk'} \cap d_{mk} - s_{mk} > d_{mk'} - s_{mk'}) \\ &= \int_{s_{mk}}^{h_k} \int_{s_{mk'}}^{d_{mk} - s_{mk} + s_{mk'}} f(d_{mk}) f(d_{mk'}) dd_{mk} dd_{mk'} \end{aligned}$$

Summing up the two probabilities yields the overall transfer failure rate:

$$P_m^{kk'} = \int_{s_{mk}}^{h_k} \int_0^{s_{mk'}} f(d_{mk}) f(d_{mk'}) dd_{mk} dd_{mk'} + \int_{s_{mk}}^{h_k} \int_{s_{mk'}}^{d_{mk} - s_{mk} + s_{mk'}} f(d_{mk}) f(d_{mk'}) dd_{mk} dd_{mk'}$$

Reference

- Bookbinder, J.H., and Desilets, A., 1992. Transfer optimization in a transit network. *Transportation Science*, 26(2), 106-118.
- Randolph, W.H., 2001. Vehicle scheduling at a transportation terminal with random delay enroute. *Transportation Science*, 19(3), 308-320.
- Ceder, A., Golang, B., Tal, O., 2001. Creating bus timetables with maximal synchronization. *Transportation Research Part A*, 35(10), 913-928.
- Chen, H., Gu, W., Cassidy, M.J., Daganzo, C.F., 2015. Optimal transit service atop ring-radial and grid street networks: A continuum approximation design method and comparisons. *Transportation Research Part B*, 83(3), 755-774.
- Chowdhury, M.S., Chien, S.I.J., 2001. Dynamic vehicle dispatching at the intermodal transfer station. *Transportation Research Record*, 1753, 61-68.
- Chowdhury, M.S., Chien, S.I.J., 2011. Joint optimization of bus size, headway, and slack time for efficient timed transfer. *Transportation Research Record*, 2218, 48-58.
- Daganzo, C.F., Pilachowski, J., 2011. Reducing bunching with bus-to-bus cooperation. *Transportation Research, Part B*, 45, 267-277.
- De Cea, J., Fernández E. Transit assignment for congested public transport systems: an equilibrium model. *Transportation Science*, 1993; 27(2): 133-147.
- Dessouky, M., Hall, R., Nowroozi A., Mourikas K., 1999. Bus dispatching at a timed transfer transit stations using bus tracking technology. *Transportation Research Part C*, 5(4), 187-208.
- Dessouky, M., Hall R., Zhang, L., Singh, A., 2003. Real-time control of buses for schedule coordination at terminal.

- Transportation Research Part A, 37(2), 145-164.
- Eranki, A., 2004. *A Model to Create Bus Timetables to Attain Maximum Synchronization Considering Waiting Times at Transfer Stops*. Master's Thesis, University of South Florida.
- Gschwender, A., Jara-Díaz, S., Bravo, C., 2016. Feeder-trunk or direct lines? Economies of density, transfer costs and transit structure in an urban context. *Transportation Research Part A*, 88, 209-222.
- Hadas, Y., Ceder, A., 2010. Optimal coordination of public-transit vehicles using operational tactics examined by simulation. *Transportation Research Part C*, 18, 879-895.
- Huang, D., Liu, Z., Liu, P., Chen, J., 2016. Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure. *Transportation Research Part E*, 96, 1-19.
- Ibarra-Rojas, O.J, Rios-Solis, Y.A., 2012. Synchronization of bus timetabling. *Transportation Research Part B*, 46(5), 599-614.
- Ibarra-Rojas, O.J, Muñoz, J.C., 2016. Synchronizing different transit lines at common stops considering travel time variability along the day. *Transportmetrica A*, 12(8), 751-769.
- Ibarra-Rojas, O.J., Loprz-Irarragorri F., Rios-Solis, Y.A., 2015. Multiperiod bus timetable. *Transportation Science*, doi: <http://dx.doi.org/10.1287/trsc.2014.0578>
- Ibarra-Rojas, O.J., Delgado, F., Giesen, R., Munoz, J.C., 2015. Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B*, 77, 38-75.
- Kim, M., Schonfeld, P., Integration of conventional and flexible bus services with timed transfers. *Transportation Research Part B*, 68, 76-97.
- Kim J., et al., 2017, Route choice stickiness of public transport passengers: Measuring habitual bus ridership behaviour using smart card data. *Transportation Research Part C*, 83,146-164.
- Liu, R., Sinha, S. Modelling urban bus service and passenger reliability. Paper presented at the International Symposium on Transportation Network Reliability, The Hague, July 2007.
- Liu T., Ceder, A., Ma J., Guan W., 2014. Synchronizing public transport transfers by using intervehicle communication scheme. *Transportation Research Record*, 2417, 78-91.
- Liu T., Ceder, A., 2017. 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*, in press.
- Marguier, P.H., Ceder, A., 1984. Passenger waiting strategies for overlapping bus routes. *Transportation Science*, 18(3), 207-230.
- Mpccia, L., Laporte, G., 2016. Improved models for technology choice in a transit corridor with fixed demand. *Transportation Research Part B*, 83, 245-270.
- Nesheli M.M, Ceder A. 2014. Optimal combinations of selected tactics for public-transport transfer synchronization. *Transportation Research Part C*, 48, 491-502.
- Parbo, J., Nielsen, O.A., Giacomo, C., 2014. User perspectives in public transport timetable optimisation. *Transportation Research Part C*, 48, 269-284.
- Poon, M.H., Wong, S.C., Tong, C.O., 2004. A dynamic schedule-based model for congested transit networks. *Transportation Research Part B*, 38(4), 343-368.
- Osuna, E.E., Newell, G.F., 1972. Control strategies for an idealized public transportation system. *Transportation Science*, 6(1), 52-72.
- Randolph, W.H., 1985. Vehicle scheduling at a transportation terminal with random delay enroute. *Transportation Science*, 19(3), 308-320.
- Si, B., Fu, L., Liu, J., Shiravi, S., Gao, Z., 2016, A multi-class transit assignment model for estimating transit passenger flows-a case study of Beijing subway network. *Journal of Advanced Transportation*, 50(1), 50-68.

- Sorratini, J.A., Liu, R., Sinha, S. Accessing bus transport reliability using micro-simulation. *Transportation Planning and Technology*, Vol.32, No.3, 2008, pp.303-324.
- Szeto, W.Y., Jiang, Y., 2014. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. *Transportation Research Part B*, 67, 235-263.
- Sadi, S., Wirasinghe S.C., Kattan Lina, 2016. Long-term planning for ring-radial urban rail transit networks. *Transportation Research Part B*, 86, 128-146.
- Schmöcker, J.D., Sun, W., Fonzone, A., Liu, R., 2016. Bus bunching along a corridor served by two lines. *Transportation Research Part B*, 93, 300-317.
- Sivakumaran, K., Li, Y., Cassidy, M.J., Madanat, S., 2012. Cost-saving properties of schedule coordination in a simple trunk-and-feeder transit system. *Transportation Research Part B*, 46, 131-139.
- Ting, C.J, Schonfeld, P., 2005. Schedule coordination in a multiple hub transit network. *Journal of Urban Planning and Development*, 131(2), 112-124.
- Ting, C.J, Schonfeld, P., 2007. Dispatching control at transfer stations in multi-hub transit networks. *Journal of Advanced Transportation*, 41(3), 217-243.
- Ting, C.J, Schonfeld, P., 2007. Dispatching control at transfer stations in multi-hub transit networks. *Journal of Advanced Transportation*, 41(3), 217-243.
- Wong, R.C.W, Yuen, T.W.Y., Fung, K.W, Leung, J.M.Y. 2008. Optimizing Timetable Synchronization for Rail Mass Transit. *Transportation Science* 42(1), 57-69.
- Wu, W., Liu, R. Jin, W., 2016. Designing robust schedule coordination scheme for transit networks with safety control margins. *Transportation Research Part B*, 93, 495-519.
- Wu, W., Liu, R., Jin, W., 2017. Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour. *Transportation Research Part B*, 104, 175-197.
- Wu, Y., Yang, H., Tang, J., Yu, Y., 2016. Multi-objective re-synchronizing of bus timetable: Model, complexity and solution. *Transportation Research Part C*, 67, 149-168.
- Wu, Y., Tang, J., Yu, Y., Pan, Z., 2015. A stochastic optimization model for transit network timetable design to mitigate the randomness of traveling time by adding slack time. *Transportation Research Part C*, 52, 15-31.
- Xu, H., Yang, H., Zhou, J., Yin, Y., 2017a. A route choice model with context-dependent value of time. *Transportation Science*, 51(2), 536-548.
- Xu, W., Zhang, W., Li, L., 2017b. Measuring the expected locational accessibility of urban transit network for commuting trips. *Transportation Research Part D*, 51, 62-81.
- Yao, J., Shi, F., Zhou, Z., Qin J., 2012. Combinatorial Optimization of exclusive bus lanes and bus frequencies in multi-modal transportation network. *ASCE Journal of Transportation Engineering*, 138(12), 1422-1429.
- Yu, B., Lu, K., Yao, S., Yao, B., Gao, Z., 2015. A bi-level programming for bus lane network design. *Transportation Research Part C*, 55, 310-327.
- Yu, B., Wu, S., Yao, B., Yang, Z., Sun, J., 2012. Dynamic vehicle dispatching at a transfer station in public transportation system. *Journal of Transportation Engineering*, 138(2), 191-201.
- Zhao, J., Dessouly, M., Bukkapatnam, S., 2006. Optimal slack time for schedule-based transit operations. *Transportation Science*, 40(4), 529-539.
- Zhao, X., Wan, C., Sun, H., Xie, D., Gao, Z., 2017. Dynamic rerouting behavior and its impact on dynamic traffic patterns. *IEEE Transactions on Intelligent Transportation System*, 99,1-17.