Transportation Research Part B 93 (2016) 181–206

Contents lists available at ScienceDirect

Transportation Research Part B

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

Reliability-based stochastic transit assignment: Formulations and capacity paradox

Y. Jiang[a,b], W.Y. Szeto[a,b,*]

[a] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, PR China
[b] The University of Hong Kong Shenzhen Institute of Research and Innovation Shenzhen, China

ARTICLE INFO

Article history:
Received 26 May 2015
Revised 15 April 2016
Accepted 21 June 2016

Keywords:
Frequency-based transit assignment
Reliability-based user equilibrium
Variational inequality
Braess and capacity paradox

ABSTRACT

This study develops link-based and approach-based variational inequality (VI) formulations for the frequency-based transit assignment with supply uncertainty, where link flows and flow on each outgoing link from each node are decision variables, respectively. Both the mean and variance of travel cost, including the covariance of in-vehicle travel costs, are captured in both formulations. To address the covariance of in-vehicle travel costs between different links on the same transit line, an augmented route-section network representation is developed, allowing us to apply the dynamic programming method to compute the value of the mapping function of the VI. The approach-based formulation can be solved by an extragradient method that only requires mild assumptions for convergence. It is found that the number of links carrying flow and equilibrium cost can be underestimated if supply uncertainty is not considered.

The study also introduces and examines the capacity paradox, a phenomenon in which the network maximum throughput may be reduced after new transit lines are added to a transit network or after the frequency of an existing line is increased. It is found that the capacity paradox may or may not occur simultaneously with the Braess-like paradox, a phenomenon in which providing new transit lines to a network may deteriorate the network performance in terms of the total weighted sum of the mean and variance of travel cost of all of the passengers. The demand level and the degree of risk aversion of passengers are the key factors that determine the occurrence of the capacity paradox.

1. Introduction

1.1. Motivations and objectives

This research was motivated by two problems. First, due to various factors such as road incidents, signal breakdown, and weather conditions, the cost components of transit assignment problems are stochastic. Although some studies (e.g., Yang and Lam, 2006; Li et al., 2008, 2009b; Sumalee et al., 2011; Meng and Qu, 2013; Szeto et al., 2013; Fu et al., 2014) have developed models to capture the stochastic costs of transit assignment problems, these models have some of the following drawbacks: (1) their formulations require specific travel time distributions, which may not be validated in reality; (2) their

* Corresponding author.
E-mail address: ceszeto@hku.hk (W.Y. Szeto).

http://dx.doi.org/10.1016/j.trb.2016.06.008

algorithms require strong conditions (e.g., monotonicity) for convergence, which may not be satisfied by the cost function of the problem; and (3) their formulations (e.g., Szeto et al., 2013) are path-based. A traditional drawback of path-based models has been the large memory and storage requirements of path enumeration. Recently, efficient methods, such as event dominance (Florian, 1998, 2004) or equilibrated choice sets (Watling et al., 2015; Rasmussen et al., 2015), have been developed to overcome these issues and have been applied to commercial software packages (e.g., Emme). However, to the best of our knowledge, most existing path set generation methods assume deterministic travel costs, and the variance and covariance of travel costs are not addressed. These methods are sometimes heuristic and cannot be easily extended to capture the covariance terms in the variance path travel cost. Szeto et al. (2011, 2013) used a *k*-shortest path algorithm to generate paths as needed; however, from the perspective of efficiency, it is hard to determine a good choice for the value of *k* in advance and they did not demonstrate their solution methods using a large, realistic network.

Our second motivation is related to the use of transit assignment models to evaluate the effectiveness of network design strategies, such as adjustments to transit itineraries or service frequencies. Without the consideration of the demand and supply uncertainties, Szeto and Jiang (2014) revealed a Braess-like paradox, in which the system performance may deteriorate, in terms of the expected total system travel cost, after new transit lines are provided to a transit network or after the frequency of an existing transit line is increased. Other than the total system travel cost, which is a level of service measure, the network performance can also be measured by the network capacity, defined as the maximum throughput of a network, which determines whether the transit network can handle all of the demand. Thus, it is necessary to investigate how transit network design strategies affect network capacity, and whether there is a paradox from the perspective of transit network capacity; that is, the network capacity can be worse off after new transit lines are provided to a transit network or after the frequency of an existing transit line is increased, i.e., the capacity paradox.

The objectives of this study are as follows.

- To develop formulation approaches to model the transit assignment problem, in which stochasticities in the in-vehicle travel cost, dwell cost, and congestion cost caused by supply uncertainty are considered. To formulate the problem, these approaches only require the mean and variance of these cost components (and the covariance of in-vehicle travel costs) without specifying their distributions.
- To analyze the properties of the problem.
- To develop a convergent solution method under milder conditions that does not rely on path enumeration and generation techniques, and illustrate its performance to solve large transit networks.
- To introduce and examine the capacity paradox.

1.2. Literature review

Some of the early works in transit assignment or related areas, such as Lampkin and Saalmans (1967), Dial (1967), and Fearnside and Draper (1971), computed the shortest path and assigned passengers on it after accounting for waiting times at transit stops. However, these early models did not consider the route choice problem of passengers traveling between a pair of stops served by several competing, direct lines, where some of the routes may be overlapped.

Le Clercq (1972) did consider the route choice problem, but assumed that passengers consider all of the direct lines and board the first arriving bus. Chriqui (1974) and Chriqui and Robillard (1975) assumed that a passenger only considers a subset of these direct lines, so as to minimize his or her expected travel time. They solved the problem of selecting the optimal subset of direct lines analytically, a problem referred to as the common line problem.

The idea of a set of attractive lines has been generalized to the optimal strategy concept (Spiess, 1984; Spiess and Florian, 1989). Assuming that a passenger will use his or her individual optimal strategy in traveling, Spiess and Florian (1989) developed a linear programming model to tackle the common line problem; their proof demonstrated that their model's dual solutions satisfy the user equilibrium conditions. Subsequently, two modeling streams were derived from the abovementioned behavioral assumption using two different network representations: the hyperpath graph network representation (Nguyen and Pallottino, 1988; Wu et al., 1994; Cominetti and Correa, 2001; Cortés et al., 2013; Sun et al., 2013) and the route-section network representation (de Cea and Fernández, 1993; Lam et al., 1999, 2002; Li et al., 2009b; Szeto et al., 2011, 2013). Although both of these modeling approaches are based on the same behavioral assumption, they have different pros and cons. The merit of the hyperpath graph representation is that the optimal set of attractive lines can be easily determined, but at the cost of creating more boarding and alighting nodes. The route-section representation can reduce the number of links required to form the network when the number of common lines is large. Moreover, the route-section representation allows the development of a link-based formulation and the adoption of available algorithms to solve for solutions.

Other network representations were developed from the hyperpath or route-section network representations, including: (a) state augmented network (Lo et al., 2003, 2004; Lozano and Storchi, 2001); (b) space time network (Nguyen et al., 2001; Hamdouch and Lawphongpanich, 2008; Hamdouch et al., 2014); (c) diachronic network (Nuzzolo et al., 2001; Sumalee et al., 2009); and (d) star network (Tong and Wong, 1999; Zhang et al., 2010). In general, a more complicated network representation requires more memory storage and computation time; however, it captures more of the cost components considered by passengers, such as non-linear transit fare, transfer cost, and congestion cost.

Among these cost components, the congestion cost due to capacity constraints has received the most attention in the literature. Existing studies adopt two approaches to address capacity constraints: the soft capacity constraint approach (e.g., Spiess and Florian, 1989; de Cea and Fernández, 1993; Nielsen, 2000, 2004; Cominetti and Correa, 2001; Lam et al., 2002; Lo et al., 2002; Nielsen and Frederiksen, 2006; Teklu 2008; Ren et al., 2009; Leurent, 2012; Liu and Meng, 2014) and the hard capacity constraint approach (e.g., Last and Leak, 1976; Lam et al., 1999; Cepeda et al., 2006; Li et al., 2009a; Szeto et al., 2013; Chen et al., 2015). The difference is that the soft capacity constraint approach allows a passenger to board a fully occupied transit vehicle, whereas the hard capacity constraint approach does not. The hard capacity constraint approach is more realistic in some situations (e.g., the high-frequency minibus service in Hong Kong where no standing is allowed). However, there may be no solution to the problems that result from insufficient capacity. The advantage of the soft capacity constraint approach is that a solution must exist to the resultant transit assignment problem under the condition that the solution set is compact and convex and the mapping function is continuous (see, e.g., theorem 1.4 of Nagurney, 1993; Szeto et al., 2013), but there is a risk of generating unrealistic link flows that exceed the link capacities.

Most of the preceding transit assignment models consider the concept of user equilibrium; that is, passengers are assumed to use the shortest path formed by a sequence of transfer stops. However, these models assume that the passengers have perfect knowledge about the network condition, which may not be realistic. Some researchers (e.g., Lam et al., 1999; 2002) extended the consideration to logit-based stochastic user equilibrium. However, they did not consider the path overlapping issue. To address this issue, researchers (Hoogendoorn-Lanser et al., 2005; Hoogendoorn-Lanser, 2005; Anderson, 2013; Anderson, et al., 2014) developed path-size logit models and probit-based models (Nielsen, 2000, 2004; Nielsen and Frederiksen, 2006) for transit assignment.

Most of the abovementioned studies assume that the cost components are deterministic, and thus ignore the variability in the travel cost components. However, the cost components are stochastic because of road incidents, signal breakdown, and weather conditions. Indeed, various empirical studies have demonstrated that passengers' route choice behavior is affected by variations in trip time caused by the supply uncertainty of transit networks (Chen et al., 2009; Casello et al., 2009; Habib et al., 2011; Frumin and Zhao, 2012; Carrel et al., 2013; Diab and El-Geneidy, 2013). Although some studies (e.g., Yang and Lam, 2006; Li et al., 2008, 2009b; Sumalee et al., 2011; Meng and Qu, 2013; Szeto et al., 2013; Fu et al., 2014) have developed stochastic models to capture the effects of supply uncertainty, they suffer from the drawbacks mentioned in Section 1.1. These issues are addressed by the first three objectives of this study.

The third objective, to develop a convergent solution method under milder conditions, is also motivated by the desire to improve the solution methods for transit assignment models even without considering uncertainty. Existing models are often solved by methods that require strong conditions to be satisfied to guarantee convergence. For example, the symmetric linear method (e.g., Wu et al., 1994) requires that the link cost function is strictly monotone[1] for convergence. The methods of de Cea and Fernández (1993) (i.e., the diagonalization method) and of Szeto et al. (2013) (i.e., self-adaptive projection and contraction algorithm with column generation) assumed monotonic[2] mapping to ensure convergence. Kurauchi et al. (2003), Cepeda et al. (2006), Sumalee et al. (2009), Schmöcker et al. (2011), and Cortés et al. (2013) adopted the MSA, whose convergence requires the cost or mapping function to be strictly pseudo-contractive (Johnson, 1972) or to satisfy the assumptions in Blum's theorem (Blum, 1954). However, these conditions are not always satisfied, especially when asymmetric link cost functions are used in transit assignment. Szeto and Jiang (2014) used the extragradient method to solve their transit assignment problem. This method only requires the mapping function to satisfy the pseudomonotone[3] and Lipschitz continuity[4] conditions for convergence. However, this method cannot be used to handle our problem directly. Liu et al. (2009) developed the self-regulated averaging method (SAM), which includes the method of successive averages (MSA) as a special case, to solve stochastic user equilibrium assignment problems. The SAM has the same convergence requirements as the MSA, but was shown to have a faster convergence speed than the MSA. The SAM has not been used to solve our studied problem.

The final objective is actually motivated by the Braess paradox in traffic assignment, the Braess-like paradox in transit assignment (Szeto and Jiang, 2014), and the capacity paradox in traffic assignment, as introduced by Yang and Bell (1998).

1.3. Contributions and organization

The contributions of this research are as follows:

- proposes link- and approach-based equilibrium conditions for transit assignment that capture supply uncertainty and the corresponding VI formulations for the problem;
- develops an augmented route-section network representation to internalize the covariance of the in-vehicle travel costs between different links on the same transit line and prevent irrational transfers, where a passenger alights from one line and boards back onto the same line;
- proves that the mapping function in the link-based VI formulation is not strictly monotone and multiple solutions exist to the problem;

[1] A vector function $\boldsymbol{F}$ is strictly monotone on a non-empty set C if for all $\mathbf{x}, \mathbf{y} \in C$, $\mathbf{x} \neq \mathbf{y}$, $(\mathbf{y}-\mathbf{x})^T(\mathbf{F}(\mathbf{y})-\mathbf{F}(\mathbf{x})) > 0$.

[2] A vector function $\boldsymbol{F}$ is monotone on a non-empty set C if for all $\mathbf{x}, \mathbf{y} \in C$, $\mathbf{x} \neq \mathbf{y}$, $(\mathbf{y}-\mathbf{x})^T(\mathbf{F}(\mathbf{y})-\mathbf{F}(\mathbf{x})) \geq 0$.

[3] A vector function $\boldsymbol{F}$ is pseudomonotone on a non-empty set C if for all $\mathbf{x}, \mathbf{y} \in C$, $(\mathbf{y}-\mathbf{x})^T\mathbf{F}(\mathbf{x}) \geq 0$ implies that $(\mathbf{y}-\mathbf{x})^T\mathbf{F}(\mathbf{y}) \geq 0$.

[4] A vector function $\boldsymbol{F}$ is Lipschitz continuous with Lipschitz constant L_f on a subset I of $\mathbb{R}^n$ if there is a nonnegative constant L_f such that $|\mathbf{F}(\mathbf{x})-\mathbf{f}(\mathbf{y})| \leq L_f|\mathbf{x}-\mathbf{y}|, \forall \mathbf{x}, \mathbf{y} \in I$.

- shows that the SAM may not obtain a convergent solution to the link-based formulation;
- illustrates the existence of a capacity paradox in transit assignment in addition to a Braess-like paradox under supply uncertainty;
- identifies the factors that affect the occurrence of the capacity paradox, and illustrates their effects with examples; and
- demonstrates that the capacity paradox may not occur simultaneously with the Braess-like paradox, indicating a need to solve a bi-objective bilevel network design problem.

In terms of formulation, although this study extends the link-based formulation and approach-based formulation introduced by Szeto and Jiang (2014) to capture supply uncertainty, the extensions are not straightforward due to the need to consider the covariance of link travel costs in the variance path travel cost. Variance path travel cost is not simply equal to the sum of the variance of link travel costs associated with the path; it also includes the covariance travel costs between links on the path. This non-additive property of variance path travel cost can be captured easily if a path-based formulation for the problem with soft (e.g., Szeto et al., 2011) or hard capacity constraints (Szeto et al., 2013) is used, because paths are clearly defined in the formulation. However, paths are not used in formulating link-based and approach-based transit assignment problems. This prevents us from directly capturing the covariance of link travel costs in the link-based and approach-based formulation approaches. To address this issue, we introduce an augmented route-section network representation that prevents irrational transfers, formulate the studied transit assignment problem based on this presentation, and mathematically prove that the covariance terms can be internalized into the resultant formulations under a realistic assumption that passengers do not make irrational transfers.

The formulation approach actually determines the choice of solution method, which might affect convergence and computation speed. Based on the approach-based formulation, we can solve the studied transit assignment problem without relying on column (i.e., path) generation heuristics or using a time-consuming path enumeration process to determine paths, which may be an advantage for two reasons. First, to the best of our knowledge, although smarter column generation techniques exist, most existing path set generation methods assume deterministic travel costs, and thus the variance and covariance of travel costs are not addressed. These methods are sometimes heuristic and cannot be easily extended to capture the covariance terms that occur in variance path travel cost. Second, although Szeto et al. (2011, 2013) used a k-shortest path algorithm to generate paths when needed, from the perspective of efficiency, a good choice of the value of k is hard to determine in advance. Szeto et al. did not demonstrate their solution methods using a large, realistic network. In contrast, the approach-based formulation allows us to solve the resulting problem (with soft capacity constraints) using the extragradient method, which guarantees convergence under milder conditions than the solution method developed by Szeto et al. (2011). The extragradient method only requires pseudomonotonicity and Lipschitz continuity of the mapping function for convergence, whereas the method of Szeto et al. (2011) requires monotonicity of the mapping function. This study proves that monotonicity is not satisfied at the level of link effective travel cost. The solution approach is demonstrated using the Winnipeg network, which consists of 1067 nodes, 3647 origin destination pairs, and 133 transit lines.

In contrast to the path-based formulations introduced by Szeto et al. (2011, 2013), this study develops link-based and approach-based formulations for the reliability-based transit assignment problem with soft capacity constraints; it adopts the extragradient method, which allows convergence under milder assumptions and avoids using column generation heuristics and the time-consuming path enumeration approach. The performance of the solution algorithm is demonstrated using a large network. Unlike Szeto and Jiang (2014), this study captures the supply uncertainty, which is not a straightforward extension due to the consideration of covariance between different sections. Unlike the literature, this study introduces and analyzes the capacity paradox, develops a new network representation that avoids irrational transfers, and proves that the travel cost is not monotone.

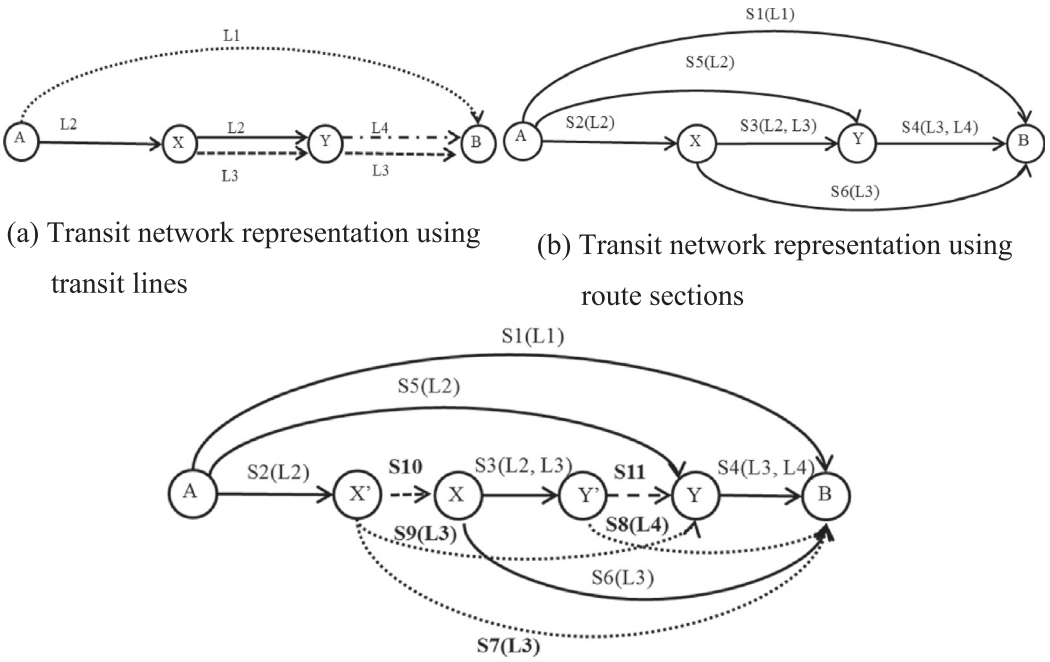
The remainder of this paper proceeds as follows. Section 2 introduces the augmented route-section network representation for a transit network, followed by the notation and assumptions used in this paper. The section then presents the path-, link-, and approach-based VI formulations. Section 3 discusses the solution algorithm for solving the approach-based VI formulation. Section 4 illustrates and analyzes the occurrence of the capacity paradox, and demonstrates the computation performance. Finally, Section 5 presents our conclusions and discusses future research directions.

2. Formulation

This section first introduces the new augmented route-section network representation, followed by our notation, assumptions, formulation of link cost components, and definition of effective travel cost. Path-based and link-based formulations are then presented for the sake of completeness. Finally, the approach-based formulation is derived.

2.1. Network representations

A general transit network is considered. The network consists of a set of transit lines and a set of transit stops (nodes), where passengers can board, alight, or transfer. The transit network can be formed using the route information provided by operators. For example, Fig. 1(a) illustrates a transit network using the lines' information provided in Table 1. For modeling purposes, the transit network is further transformed into a route-section network representation.



(a) Transit network representation using transit lines

(b) Transit network representation using route sections

(c) Augmented transit network representation using route sections

Fig. 1. Transit network representations.

Table 1
Transit line information.

Line	Stops
L1	A, B
L2	A, X, Y
L3	X, Y , B
L4	Y, B

2.1.1. Route-section network representation

Fig. 1(b) shows the route-section network that corresponds to the line-node network presented in Fig. 1(a). The route-section network was introduced by de Cea and Fernández (1993). The aim of the route-section network representation is to classify passengers waiting at transit stops (including origins or transfers) into different groups based on their next alighting nodes (which may be their next transfer locations or their destinations). For passengers boarding at the same transit stop and traveling to the identical alighting node without a transfer, a *link* called a *section* is created to connect the boarding and alighting stops. On each section, if there is more than one line, a set of attractive lines is determined via the method introduced by Chriqui and Robillard (1975). Afterwards, the flow on an attractive bus line is obtained by splitting the section flow according to the line's relative frequency.

The transit assignment over a traditional route section network may overestimate the number of transfers and generate irrational transfers within the same line. For example, if 100 passengers travel from A to Y in Fig. 1(b) and if 60 passengers select the route that passes section S5, then the remaining 40 passengers select the route that passes S2(L2) and S3(L2, L3). Assuming that the frequencies of L2 and L3 are identical, the 40 passengers traveling on S3(L2, L3) are equally assigned to L2 and L3. This result means that on S2, there are 20 passengers on L2 and 20 passengers on L3. More precisely, the results imply that 20 passengers board L2 at node A, alight at node X, then board L2 again. Such transfers are rarely found in reality, as passengers can travel to node Y directly via L2. Thus, it should be eliminated from the results.

2.1.2. Augmented route-section network representation

To address this issue, a new augmented route-section network representation is proposed, as shown in Fig. 1(c). The aim is to create a dummy (route) section with a large travel cost to penalize transfers within the same line; at the same time, a dummy node is created to allow transfers between different lines. For example, in Fig. 1(c), dummy sections S10 and S11, represented by dashed lines, are created to connect X′ to X and Y′ to Y, respectively, where X′ and Y′ are dummy nodes representing transfer stops between different lines. In addition, new sections, denoted by dotted lines, S7(L3), S8(L4), and

S9(L3), are constructed to model transfers between different lines. For example, if passengers using S2(L2) want to transfer to L3, they can alight at dummy node X′ and transfer to the new sections S7(L3) and S9(L3) to arrive at nodes B and Y, respectively. Similarly, passengers using S3(L2, L3) can transfer to L4 at node Y′

On the augmented network, any route containing an irrational transfer must pass at least one dummy section, and any route not containing any irrational transfer must not pass any dummy section. For example, if passengers want to travel from A to Y via S2(L2) and S3(L2, L3), they must pass dummy section S10. Because the travel cost on S10 is high, no passengers are assigned to route S2(L2)-S3(L2, S3) and irrational transfers are avoided. In fact, passengers boarding L2 can directly arrive at node Y, represented by S5(L2), or transfer to L3 at X′, represented by S9(L3). In these two cases, no dummy sections and no irrational transfers are involved.

As shown in Appendix A, in addition to eliminating the routes containing irrational transfers, using the augmented route-section network representation under the assumption that passengers do not make irrational transfers allows the internalization of the covariance of the in-vehicle travel costs between different sections of the same line into the variance of one section. Thus, the resultant problem can be solved as a traditional transit assignment problem without considering supply uncertainty; thus, the dynamic programming method can be applied to solve the resultant formulation. Note that it is important to model the covariance between different sections of the same line, as the in-vehicle travel costs associated with the same line on different sections are dependent. For example, a delay on one section can lead to a delay on a subsequent section of the same line in reality.

Although the augmented route-section network representation has advantages for handling covariance and rational and irrational transfers properly, it requires more computational time and storage due to the additional dummy links and dummy nodes created. Therefore, the choice to use the augmented network depends on the trade-off between the requirements of modeling accuracy and computational efficiency.

In addition, it is worth mentioning that the augmented route-section network representation was initially motivated by the need for link-based and approach-based formulations. There is no conceptual difficulty in applying it to a path-based formulation. Nevertheless, if a path-based formulation is developed, irrational transfers could be alternately addressed by event-dominance principles (Florian, 2004 and Nielsen, 2006) or transfer penalties (Nuzzolo et al., 2012).

2.2. Notation

2.2.1. Set

N	set of nodes or stops
S	set of sections, links, or approaches
O	set of origins
D	set of destinations
Q	set of origin-destination (OD) pairs
$L_s, \bar{L}_s$	set of lines and attractive lines on section s, respectively
A_i^+, A_i^-	set of sections coming out from and going into node i, respectively
P^{od}	set of paths connecting an OD pair $(o, d) \in Q$
Ω_f	solution space in terms of path flows
Ω_v	solution space in terms of link flows
Ω_α	solution space in terms of approach proportions

2.2.2. Indices

r	origin index
d	destination index
s	section index
l	line index
p	path index
$t(s)$, $h(s)$	tail and head nodes of section s, where $s = (t(s), h(s))$

2.2.3. Parameters

$b_{s,1}, b_{s,2}, b_{s,3}, n$	calibrated perceived congestion parameters for section s
γ_k	unit conversion parameter for capacity; when the unit of headway is minutes and that of bus capacity is passengers per hour, $\gamma_k = 60$ minutes per hour
γ_f	unit conversion parameter for frequency; when the unit of frequency is vehicles per hour and that of waiting is minutes, $\gamma_f = 60$ minutes per hour
ρ	parameter representing the risk-aversion of passengers
f^l	mean frequency of line l
k^l	capacity of a single vehicle traversing route l
g^{od}	demand of an OD pair od or demand between nodes o and d
μ_T, μ_W, μ_Φ	value of in-vehicle travel time, value of waiting time, and value of congestion, respectively

2.2.4. Variables and functions

C_s	travel cost associated with section s
C^{id}	travel cost between nodes i and d
C_p^{od}	travel cost from origin o to destination d via path p
T_s^l	in-vehicle travel time of line l on section s
T_s	weighted average in-vehicle travel time on section s
$T_{s,0}^l$	dwell time of line l on section s
$T_{s,0}$	weighted average dwell time on section s
W_s	waiting time of passengers for the first arriving transit vehicle on section s
w_s^l	relative frequency of line l on section s
Φ_s	perceived congestion cost on section s because of insufficient capacity
f_s	mean frequency of section s
K_s	capacity of section s
$E[\cdot]$	expectation of a random variable
$Var[\cdot]$	variance of a random variable
$Cov[\cdot,\cdot]$	covariance between two random variables
v_s	number of passengers on section s
$\hat{v}_s$	total number of passengers on the competing sections of section s
v_{sl}^d	number of passengers on section s using line l in the direction of destination d
v_s^d	number of passengers on section s moving toward destination d
$\mathbf{v}$	vector of (v_s^d) with dimension $\lvert S\rvert \times \lvert D\rvert$
Δ_p^s	element in the path-section incidence matrix; if route p passes section s, $\Delta_p^s = 1$, otherwise $\Delta_p^s = 0$
f_p^{od}	flow on path p connecting OD pair od
$\mathbf{f}$	vector of (f_p^{od}) with dimension $\sum_{(o,d)\in Q} \lvert P^{od}\rvert$
u_s^{id}	minimum effective travel cost from node i to destination d via section $s \in A_i^+$
u^{id}	minimum effective travel cost between nodes i and d
$\mathbf{u}$	vector of (u_s^{id}) with dimension $\lvert N\rvert \times \lvert D\rvert$
α_s^d	proportion of passengers leaving node $t(s)$ via approach s to destination d
α	vector of (α_s^d) with dimension $\lvert S\rvert \times \lvert D\rvert$
π_p^{od}	effective travel cost associated with path p connecting OD pair od
π^{od}	minimum effective travel cost between OD pair od
q^{id}	number of passengers leaving node i toward destination d

2.3. Assumptions

As in previous studies (e.g., Spiess and Florian, 1989; de Cea and Fernández, 1993; Szeto et al., 2013; Szeto and Jiang, 2014), the following classical assumptions are made throughout this paper.

A1. The travel demand between each OD pair in the system is assumed to be known and fixed. This assumption is reasonable for strategic planning when day-to-day variation is small and negligible, and the demand for public transport is stable in the long term.

A2. Passengers are assumed to arrive at transit stops randomly.

A3. Passengers know the mean and variance of in-vehicle travel time for each section. This assumption is reasonable for frequent commuters who are well aware of the level of service.

A4. Each passenger selects the transit route that minimizes his/her effective travel cost, where the effective travel cost is defined in the next section.

A5. Each passenger waiting at a transfer node considers a set of attractive lines and boards the first arriving bus belonging to this set of attractive lines. Traditionally, the set of attractive lines can be obtained by the method developed by de Cea and Fernández (1993), in which an integer optimization model that minimizes a passenger's expected travel cost is solved to determine whether a transit line is attractive. In this study, we revise the integer optimization model using effective travel cost. For simplicity, the set of attractive lines is only determined once and assumed to be fixed despite congestion levels.

A6. The values of in-vehicle travel time, waiting time, and congestion are assumed to be independent and constant. Although it has been pointed out that these values could be random and correlated (Nielsen et al., 2002; Mabit and Nielsen, 2006), they are not considered, because the focus of this study is the stochasticity in the supply side components.

A7. Stochastic vehicle headways with the same distribution function (i.e., exponential) are assumed for vehicles servicing different lines.

A8. The in-vehicle travel costs of different lines are independent, whereas the in-vehicle travel costs associated with the same line on different sections are dependent and their covariances are known.

A9. The waiting time for a transit line on a section is independent of the waiting times for other lines on the same section.

A10. The in-vehicle travel, waiting, congestion, and dwell time costs of a section are independent of each other. This assumption requires that there is no interaction between different cost components. However, as the waiting time depends on the frequency of transit lines and the frequency is determined by the in-vehicle travel time, these cost components may not be independent. Thus, the relaxation of this assumption is left for future study.

A11. The in-vehicle travel time on a section is subject to randomness due to supply uncertainty.
A12. The waiting, congestion, and dwell costs of a section are independent of those of other sections.

2.4. Cost components

The travel cost associated with section s, C_s, is defined by

$$C_s = \mu_T T_s + \mu_W W_s + \mu_\Phi \Phi_s + \mu_T T_{s,0}, \forall s \in S. \tag{1}$$

Eq. (1) means that the travel cost on section s comprises the in-vehicle travel cost $\mu_T T_s$, the waiting cost of passengers for the first arriving vehicle $\mu_W W_s$, the perceived congestion cost $\mu_\Phi \Phi_s$ caused by insufficient vehicle capacity, and the dwell time cost $\mu_T T_{s,0}$, where $\Phi_s = b_{s,1}\left(\frac{b_{s,2}v_s + b_{s,3}\hat{v}_s}{\sum_{l\in\bar{L}_s} f^l k^l}\right)^n$. Note that the coefficient associated with a section flow is different from that associated with a competing section, because the congestion cost of a passenger due to waiting time at a stop may be higher than that due to in-vehicle congestion. Due to supply uncertainty, these cost components are stochastic and modeled as random variables. Based on the assumptions given in Section 2.3, their means and variances are derived as follows:

$$E[\mu_T T_s] = E\left[\mu_T \sum_{l\in\bar{L}_s} w_s^l T_s^l\right] = \mu_T \sum_{l\in\bar{L}_s} w_s^l E\left[T_s^l\right], \forall s \in S, \tag{2}$$

$$Var[\mu_T T_s] = \mu_T^2 \sum_{l\in\bar{L}_s} \left(w_s^l\right)^2 Var\left[T_s^l\right] + \mu_T^2 \sum_{l\in\bar{L}_s}\sum_{l'\in\bar{L}_s} Cov\left[T_s^l, T_s^{l'}\right], \forall s \in S, \tag{3}$$

$$E[W_s] = \mu_W \frac{\gamma_f}{f_s}, \forall s \in S, \tag{4}$$

$$Var[W_s] = \mu_W^2 \left(\frac{\gamma_f}{f_s}\right)^2, \forall s \in S, \tag{5}$$

$$E[\Phi_s] = \mu_\Phi b_{s,1} n! \left(\frac{b_{s,2}v_s + b_{s,3}\hat{v}_s}{\sum_{l\in\bar{L}_s} f^l k^l \gamma_k}\right)^n, \forall s \in S, \tag{6}$$

$$Var[\Phi_s] = \mu_\Phi^2 (b_{s,1})^2 \left((2n)! - (n!)^2\right)\left(\frac{b_{s,2}v_s + b_{s,3}\hat{v}_s}{\sum_{l\in\bar{L}_s} f^l k^l \gamma_k}\right)^{2n}, \forall s \in S, \tag{7}$$

$$E[T_{s,0}] = \mu_T \sum_{l\in\bar{L}_s} w_s^l E\left[T_{s,0}^l\right], \forall s \in S, \quad \text{and} \tag{8}$$

$$Var[T_{s,0}] = \mu_T^2 \sum_{l\in\bar{L}_s} \left(w_s^l\right)^2 Var\left[T_{s,0}^l\right], \forall s \in S, \tag{9}$$

where w_s^l is the relative frequency of line l on section s and defined by

$$w_s^l = \frac{f^l}{f_s}, \forall s \in S, l \in \bar{L}_s \quad \text{and} \tag{10}$$

f_s is the mean frequency associated with section s and is given by

$$f_s = \sum_{l\in\bar{L}_s} f^l, \forall s \in S. \tag{11}$$

The details of the derivations and discussions on the means and variance of W_s and Φ_s can be found in Szeto et al. (2011). The other derivations are straightforward.

2.5. Effective travel cost

The variability in the in-vehicle travel time and the waiting time, along with the delay due to congestion, cause variability in trip time (see, e.g., Lo et al., 2006) and in the value of travel time (see, e.g., Senna, 1994; Fosgerau and Engelson, 2011). Consequently, a passenger cannot determine the exact trip time/cost to complete his or her journey. The passenger counters the variability in trip time/cost by an early departure to allow for additional time for the trip and avoid being late. The additional time is commonly referred to as a safety margin, and depends on both the purpose of the trip and the individual's risk taking behavior. Lo et al. (2006), Shao et al. (2008), Lam et al. (2008), and Siu and Lo (2008) calculate the effective travel time as the safety margin plus the expected trip time. However, this concept does not consider the fact that the monetary

value of in-vehicle travel time differs from that of the waiting time. Szeto et al. (2011) proposed the concept of effective travel cost (or travel cost budget), which generalizes the concept of travel time budget by using the trip travel cost (including in-vehicle travel time cost and waiting time cost) instead of the trip travel time. The safety margin in terms of travel time or cost can be represented by a linear function of either standard deviation or variance. Most studies have adopted the former but this study adopts the latter due to its computational efficiency. Mathematically, the effective travel cost associated with route p between OD pair od, π_p^{od}, can be expressed as

$$\pi_p^{od} = E\left[C_p^{od}\right] + \rho Var\left[C_p^{od}\right], \forall (o,d) \in Q, p \in P^{od}, \tag{12}$$

where ρ is the parameter that represents the degree of risk aversion of passengers and the unit is equal to the reciprocal unit of the mean route travel time; this can be calibrated using the method in Jackson and Jucker (1982). ${C_p}^{od}$ represents the travel cost associated with route p connecting OD pair od and is defined by

$$C_p^{od} = \sum_{s \in S} \Delta_p^s C_s, \forall (o,d) \in Q, p \in P^{od}. \tag{13}$$

By substituting Eq. (13) into Eq. (12), the effective travel cost is expressed as

$$\pi_p^{od} = E\left[\sum_{s \in S} \Delta_p^s C_s\right] + \rho Var\left[\sum_{s \in S} \Delta_p^s C_s\right], \forall (o,d) \in Q, p \in P^{od}. \tag{14}$$

By substituting Eq. (1) into Eq. (14) and simplifying the resultant expression, the effective travel cost can be obtained as shown in the following equation:

$$\begin{aligned}\pi_p^{od} = & \sum_{s \in S} \Delta_p^s (E[\mu_T T_s] + E[\mu_W W_s] + E[\mu_\Phi \Phi_s] + E[\mu_T T_{s,0}]) \\ & + \rho \sum_{s \in S} \Delta_p^s (Var[\mu_T T_s] + Var[\mu_W W_s] + Var[\mu_\Phi \Phi_s] + Var[\mu_T T_{s,0}]) \\ & + \rho \sum_{s \in S} \sum_{s' \in S, s' \neq s} \Delta_p^s \Delta_p^{s'} Cov[\mu_T T_s, \mu_T T_{s'}], \forall (r,d) \in Q, p \in P^{od}.\end{aligned} \tag{15}$$

Eq. (15) assumes that each cost component is independent, and only the covariance of the in-vehicle travel costs between different links on the same transit line is non-zero.

2.6. Path-based formulation

Under the assumption that all of the passengers choose their lowest effective travel cost routes, the reliability-based user equilibrium (RUE) condition in terms of path flows is defined as follows:

$$\pi_p^{od} \begin{cases} = \pi^{od}, f_p^{od} > 0 \\ \geq \pi^{od}, f_p^{od} = 0 \end{cases}, \forall (o,d) \in Q, p \in P^{od}. \tag{16}$$

Eq. (16) indicates that if a route between OD pair od carries flow, then its effective travel cost must equal the minimum effective cost of that OD pair; otherwise, the cost is not less than the minimum effective travel cost.

The path-based RUE conditions can also be represented in a nonlinear complementarity problem (NCP) format:

$$(\pi_p^{od} - \pi^{od}) f_p^{od} = 0, \forall p \in P^{od}, (o,d) \in Q, \tag{17}$$

$$\pi_p^{od} - \pi^{od} \geq 0, \forall p \in P^{od}, (o,d) \in Q, \tag{18}$$

$$f_p^{od} \geq 0, \forall p \in P^{od}, (o,d) \in Q. \tag{19}$$

Furthermore, path flows must satisfy the flow conservation conditions:

$$\sum_{p \in P^{od}} f_p^{od} = g^{od}, \forall (o,d) \in Q. \tag{20}$$

The reliability-based transit assignment problem finds a path flow vector to satisfy (2)–(11), (14), and (16)–(19); it can be reformulated as a variational inequality (VI) problem. The VI problem determines $\mathbf{f}^* = [f_p^{od*}] \in \Omega_f$, such that

$$(\mathbf{f} - \mathbf{f}^*) \pi(\mathbf{f}^*) \geq 0, \forall \mathbf{f} \in \Omega_f, \tag{21}$$

where $\Omega_f = \{f_p^{od} | f_p^{od} \geq 0, \forall (o,d) \in Q, p \in P^{od} \text{and} \sum_{p \in P^{od}} f_p^{od} = g^{od}, \forall (o,d) \in Q\}$, $\pi(\mathbf{f}) = [\pi_p^{od}]$, and π_p^{od} is defined by (2)–(11) and (14). It is not difficult to see that the solution set is closed and compact, and the mapping function is continuous. Hence, a solution exists to the problem.

2.7. Link-based formulation

The path-based formulation requires path set information, which can be obtained by path enumeration algorithms. One method for reducing the computational burden in generating and maintaining path set is to use column generation techniques. In that case, paths are only generated when needed. The second method is to reformulate the problem into a link-based formulation, so that path enumeration is replaced by a shortest path determination procedure during the computation process. This section considers the second method.

The link-based RUE condition is defined as

$$u_s^{id}\begin{cases} = u^{id}, v_s^d > 0 \\ \geq u^{id}, v_s^d = 0 \end{cases}, \forall s \in A_i^+, i \in N, d \in D, \tag{22}$$

where u_s^{id} is the minimum effective travel cost from node i to destination d via section $s \in A_i^+$ and u^{id} is the minimum effective travel cost between nodes i and d. u_s^{id} and u^{id} are respectively defined by,

$$u_s^{id} = E\left[C_s + C^{h(s)d}\right] + \rho Var\left[C_s + C^{h(s)d}\right], \forall i \in N, d \in D, s \in A_i^+ \text{ and} \tag{23}$$

$$u^{id} = \min_{s \in A_i^+} u_s^{id}, \forall i \in N, d \in D. \tag{24}$$

If all of the travel costs components are assumed to be independent, Eq. (23) allows us to directly apply the dynamic programming approach to easily obtain solutions. However, in this study, the covariances between different sections of the same line are considered, making our model more realistic. In this case, the covariances prevent a straightforward application of the dynamic programming method. To address this issue, we introduce the realistic assumption that passengers do not transfer within the same line. Under this assumption and using the augmented route-section network, u_s^{id} can be expressed as

$$u_s^{id} = u^{h(s)d} + E[C_s] + \rho Var[C_s], \forall i \in N, s \in A_i^+, d \in D, \tag{25}$$

where $E[C_s]$ and $Var[C_s]$, respectively, equal $E[\mu_T T_s] + E[\mu_W W_s] + E[\mu_\Phi \Phi_s] + E[\mu_T T_{s,0}]$ and $Var[\mu_T T_s] + Var[\mu_W W_s] + Var[\mu_W \Phi_s] + Var[\mu_T T_{s,0}]$, and can be easily obtained using Eqs. (2) to (11). *The derivation of the above equation is presented in Appendix A.*

The link-based RUE conditions can be represented as an NCP:

$$(u_s^{id} - u^{id}) v_s^d = 0, \forall s \in A_i^+, i \in N, d \in D, \tag{26}$$

$$u_s^{id} - u^{id} \geq 0, \forall s \in A_i^+, i \in N, d \in D, \text{ and} \tag{27}$$

$$v_s^d \geq 0, \forall s \in A_i^+, i \in N, d \in D. \tag{28}$$

In addition, the flow conservation constraint is defined by

$$\sum_{s \in A_i^-} v_s^d + g^{id} = \sum_{s \in A_i^+} v_s^d, \forall i \in N, d \in D. \tag{29}$$

The link-based NCP formulation can be reformulated as a VI problem: to find $\mathbf{v}^* = [v_s^{d*}] \in \Omega_v$ such that

$$(\mathbf{v} - \mathbf{v}^*)^{\mathbf{T}} \mathbf{u}(\mathbf{v}^*) \geq 0, \forall \mathbf{v} \in \Omega_v, \tag{30}$$

where $\Omega_v = \{v_s^d | v_s^d \geq 0, \forall s \in S, d \in D \text{ and } \sum_{s \in A_i^+} v_s^d = g^{id} + \sum_{s \in A_i^-} v_s^d, \forall i \in N, d \in D\}$, $\mathbf{u}(\mathbf{v}) = [u_s^{id}]$, and the elements in $\mathbf{u}(\mathbf{v})$ can be obtained by Eqs. (2)–(11) and (24)–(25).

For the link-based VI formulation, the following propositions can be proved.

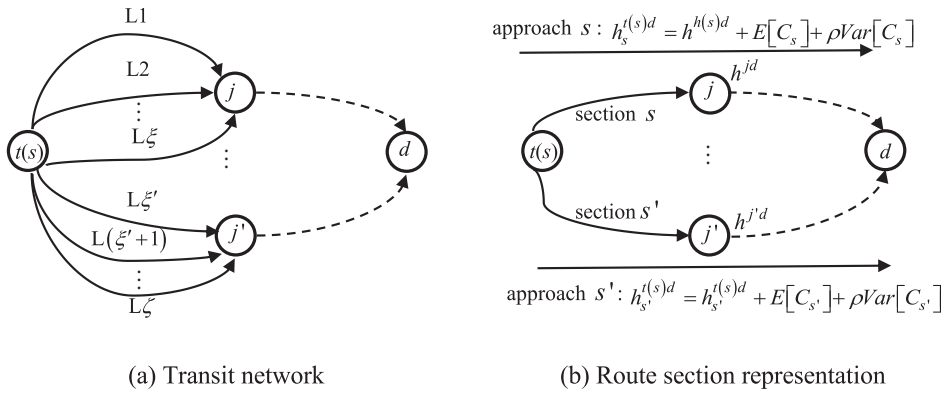
Proposition 1. *The mapping function* $\mathbf{u}(\mathbf{v})$ *in the link-based VI formulation is not strictly monotone.*

Proof: See Appendix B.

Proposition 1 implies that the solution methods that require the strictly monotone condition for convergence (i.e., Wu et al., 1994) may not give an RUE solution to the studied problem. Therefore, it is necessary to develop solution algorithms that only require mild assumptions to solve the problem. From proposition 1, we can easily derive the following proposition.

Proposition 2. *Multiple solutions exist to the link-based VI formulation.*

Proof: See Appendix B.



(a) Transit network (b) Route section representation

Fig. 2. Network representation of the approach-based formulation.

2.8. Approach-based formulation

Following Szeto and Jiang (2014), the preceding link-based formulation can be reformulated as an approach-based formulation in which the approach of a node is defined by the section *coming out* from that node, and an approach proportion is defined as the proportion of passengers leaving a node via the approach considered. If the approach proportion, α_s^d, is denoted as the proportion of the passengers per hour going to destination d via section s, the proportion must then satisfy the following conditions:

$$0 \le \alpha_s^d \le 1, \forall s \in S, d \in D \quad \text{and} \tag{31}$$

$$\sum_{s \in A_i^+} \alpha_s^d = 1, \forall i \in N, d \in D, \tag{32}$$

where A_i^+ is the set of sections coming out from node i. The approach-based RUE equilibrium conditions can be formulated as

$$(h_s^{t(s)d} - h^{t(s)d})\alpha_s^d = 0, \forall s \in A_i^+, i \in N, d \in D, \tag{33}$$

$$h_s^{t(s)d} - h^{t(s)d} \ge 0, \forall s \in A_i^+, i \in N, d \in D, \quad \text{and} \tag{34}$$

$$\alpha_s^d \ge 0, \forall s \in S, d \in D, \tag{35}$$

where $h_s^{t(s)d}$ represents the effective travel cost from the tail node of section s to destination d via approach s, and $h^{t(s)d}$ is the minimum effective travel cost between the tail node of section s and d. Note that they are functions of $\alpha = [\alpha_s^d]$.

Additionally, the following flow conservation constraint must hold:

$$\sum_{s \in A_i^-} \alpha_s^d q^{t(s)d} + g^{t(s)d} = \sum_{s \in A_i^+} \alpha_s^d q^{t(s)d}, \forall i \in N, d \in D. \tag{36}$$

The link-based transit assignment formulation in terms of approach proportions can be reformulated as a VI problem that determines $\alpha^* = [\alpha_s{}^{d*}] \in \Omega_\alpha$ such that

$$(\alpha - \alpha^*)^{\mathbf{T}} \mathbf{h}(\alpha^*) \ge 0, \forall \alpha \in \Omega_\alpha, \tag{37}$$

where $\alpha = [\alpha_s^d]$, $\mathbf{h}(\alpha) = [h_s^{t(s)d}]$, and

$$\Omega_\alpha = \left\{ \begin{matrix} \alpha_s^d | 0 \le \alpha_s^d \le 1, \forall s \in S, d \in D, \sum_{s \in A_i^+} \alpha_s^d = 1, \sum_{s \in A_i^-} \alpha_s^d q^{t(s)d} + g^{t(s)d} = \sum_{s \in A_i^+} \alpha_s^d q^{t(s)d}, \\ \forall i \in N, d \in D \end{matrix} \right\}.$$

Although we use the term “approach proportion”, the definition we use is different from that proposed by Bar-Gera (2002). We define “approach proportion” using outgoing links from nodes, whereas Bar-Gera’s definition is based on incoming links. Using outgoing links as decision variables in the definition of the proportion of flows allows the resulting formulation to easily capture the common line feature and determine how passengers select transit lines from the set of attractive lines.

The approach-based formulation is further explained in Fig. 2. Fig. 2(a) is the original transit network and Fig. 2(b) is the corresponding route-section representation. The dotted line represents the minimum effective cost path connecting

node j (j') to destination d; its cost, $h^{jd}(h^{j'd})$, is marked next to node j (j'). Assuming that L1, L2, ..., Lξ are the attractive lines connecting node $i = t(s)$ to node j, these lines are combined into one section, referred to as section s without loss of generality, with a section cost C_s. Similarly, section s' aggregates the set of attractive lines Lξ', L$(\xi'+1)$, ..., Lζ, which connect node $t(s)$ and node j'.

In Fig. 2(b), there are two paths from node $t(s)$ to destination d. One path is via section s and the other is via section s'. The former is denoted as approach s and the latter is denoted as approach s'. The proportions of these two approaches are denoted as α_s^d and $\alpha_{s'}^d$, respectively. The effective travel cost from node $t(s)$ to node d via approach s is obtained by adding the effective route-section cost $c_s = E[C_s] + \rho Var[C_s]$ to the minimum effective cost from the head node of section s (i.e., node j) to destination d. Similarly, the effective travel cost from node $t(s)$=$t(s')$ to node d via s' is obtained by adding the effective route-section cost $c_{s'} = E[C_{s'}] + \rho Var[C_{s'}]$ to the minimum effective cost from the head node of section s' (i.e., node j') to destination d.

Under the preceding supposition, if approach s carries a flow (i.e., $1 \geq \alpha_s^d > 0$), then the effective travel cost from node $t(s)$ to node d via approach s must equal the minimum effective path cost from node $t(s)$ to destination d (i.e., $h_s^{t(s)d} = h^{t(s)d}$), and approach s must be on another lowest effective cost path from node $t(s)$ to node d (i.e., $h_s^{t(s)d} = h_{s'}^{t(s)d} = h^{t(s)d}$); otherwise, conditions (33)–(35) are violated. If section s carries no flow (i.e., $\alpha_s^d = 0$) under the preceding supposition, approach s may or may not be on one of the lowest effective cost paths. Approach s is on one of these paths only if the effective cost associated with path $t(s) - j - d$ equals the effective cost associated with the lowest effective cost path $t(s) - j' - d$. Nevertheless, in most cases, the effective cost associated with path $t(s) - j - d$ is larger than that associated with $t(s) - j' - d$, and then approach s is NOT on the lowest effective cost path $t(s) - j' - d$.

3. Solution method

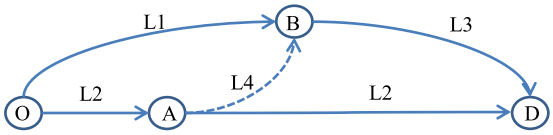
To solve the VI problem, we adopt a projection method belonging to the class of extragradient methods introduced by Korpelevich (1976). The advantage of the extragradient method is that it does not require knowing the Lipschitz constant in advance and only requires the mapping function to satisfy mild conditions, i.e., pseudomonotonicity and Lipschitz continuity, for convergence. In each iteration of this method, one projection is performed to get an approximate solution, then a second projection is performed to refine the solution. These two steps are iteratively carried out until an optimal (or acceptable) solution is found. However, this projection method cannot be directly applied to solve the transit assignment problem, because the method does not account for the specific features of the problem, such as the common line problem. Hence, Szeto and Jiang (2014) developed a revised extragradient method, in which a cost-updating algorithm was introduced. In this study, the cost-updating algorithm is further extended by incorporating the variance of travel cost into the link cost. This modification does not affect the computation complexity provided the cost components are link additive after the covariance term is internalized using the augmented route-section network. Therefore, following a procedure similar to Szeto and Jiang (2014), it can be shown that given a directed graph and the topological ordering of its nodes, the complexity of the single destination case to update expected travel cost is $O(V + E)$, where V is the total number of vertices (i.e., nodes) and E is the number of edges (i.e., route-sections or links). Due to space limitations, the details of the algorithm are not reported in this paper, but can be found in Szeto and Jiang (2014).

4. Numerical examples

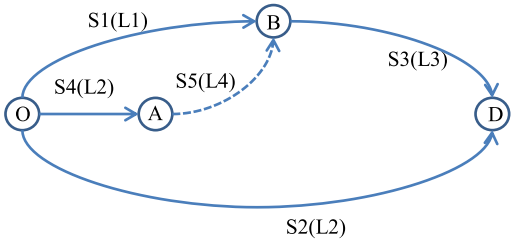
This section has six subsections. Subsection 4.6 uses the Winnipeg network to illustrate the computation performance of the developed solution method and SAM. Subsections 4.1–4.5 use the small network presented in Fig. 3 to illustrate and examine the capacity paradoxical phenomenon and the importance of considering supply uncertainty.

In Fig. 3, four lines are in service, namely, L1, L2, L3, and L4. Fig. 3(a) represents the transit network using transit line representation and Fig. 3(b) represents the network using route-section representation. Section S5(L4) is represented with a dashed line, as it is assumed that line L4 does not exist initially. Table 2(a) lists the required data. As line L2 passes two links, its mean (variance) is the sum of two numbers, where each number represents the mean (variance) of one section. For the small network, two OD pairs are considered. Table 2(b) lists the path information for each OD pair. The other parameters, unless specified otherwise, are set as $g^{OD} = 500$ and $g^{BD} = 100$ passengers/hour; $\rho = 0.05$; $k = 120$ passengers/vehicle; $\gamma_k = \gamma_f = 60$ minutes per hour; $n = 2$, $b_{s,2} = b_{s,3} = 1$, $b_{s,1} = 1.0$, $b_{s,2} = b_{s,3} = 0.1$, $b_{s,4} = 0.5$, and $b_{s,5} = 0.1, \forall s \in S$; $\mu_T = \mu_W = \mu_\Phi = 1.0$. For simplicity, the dwell time is assumed to be one minute for each section and the variance of the dwell time is assumed to be zero.

Similar to the network capacity defined by Yang and Bell (1998), the network capacity here refers to the *maximum throughput of the network at which all of the bottleneck links just reach their capacities under the reliability-based user equilibrium condition*, where a bottleneck link is the link with the lowest capacity on a path. For the ease of explanation, the network residual capacity, which is defined by the difference between the sum of the capacities on all of the bottleneck links and the flows on these bottleneck links, is presented. If the network residual capacity reduces with no change in the total demand on the network, then this indicates that the network capacity is reduced. In our case, we consider *critical sections* instead of bottleneck links, and define a critical section as the section with flow at least equal to expected section capacity.



(a) Network representation using transit lines



(b) Network representation using route sections

Fig. 3. Small network for the capacity paradox.

We want to emphasize that the network capacity is defined under the reliability-based user equilibrium condition. It is *selfish* behavior that leads to the reduction in throughput. Without considering the reliability-based user equilibrium condition, the maximum throughput can simply be determined by solving the maximum flow problem. In that case, adding a new line or increasing the frequency of an existing line will not lead to a reduction in throughput.

4.1. Occurrence of the capacity and Braess-like paradoxes

This section illustrates the occurrence of the capacity and the Braess-like paradoxes in the studied transit assignment problem, which includes the classical frequency-based transit assignment problem (e.g., de Cea and Fernández, 1993) as a special case. For the capacity paradox, the following points need to be clarified. First, unlike Yang and Bell (1998), in this study, the network capacity is the expected network capacity, as the capacity is modeled as a random variable due to the supply uncertainty. Second, the section flow is the effective section flow, which incorporates the flows on the competing sections due to the common line issue. This issue is the characteristic in transit assignment.

Table 3 presents the results of a before-after study, which reveals the capacity paradox. Before line L4 is provided, the network residual capacity is 360, meaning that the network can serve an additional 360 passengers without using up the capacities of all of the bottleneck sections. However, when line L4 is added and its frequency is set to be 9.0 buses/hour, the network residual capacity reduces to zero. Such a phenomenon is referred to as the capacity paradoxical phenomenon, as it is paradoxical that adding a new line may reduce the network maximum throughput. At the same time, the Brass-like paradox also occurs: the total effective travel cost increases from \$44,932.8 to \$45,076.9 after the new line is added. This simultaneous occurrence does not always happen; it depends on the starting and ending frequencies.

4.1.1. Causes of the paradoxes

To investigate the causes of the paradoxes and the effect of the starting and ending frequencies on the occurrence of the paradoxes, the frequency of line L4 is varied from 7.4 buses/hour to 9.4 buses/hour. The resultant network residual capacities

Table 2
Input data of the small network for the capacity paradox.

(a) Line data				(b) Path data		
Line no.	Travel time		Frequency (no. of buses/hour)	OD pair	Path No.	Section sequence
	Mean (min)	Variance (min^2)				
L1	6	3	3	(1) OD	Path 1	S1(L1)-S3(L3)
L2	5+50 *	2+3 **	4		Path 2	S2(L2)
L3	8	2	4		Path 3	S4(L2)-S5(L4)-S3(L3)
L4	3	2	–	(2) BD	Path 4	S3(L3)

* The mean travel time between nodes O and A is 5 minutes, and that between A and D is 50 minutes.
** The variance in travel time between nodes O and A is 2 min^2, and that between A and D is 3 min^2.

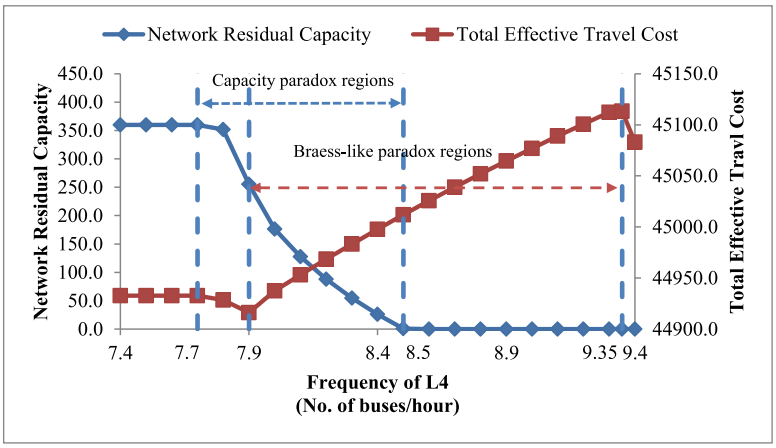


Fig. 4. Total effective travel costs and network residual capacities under various frequencies of L4.

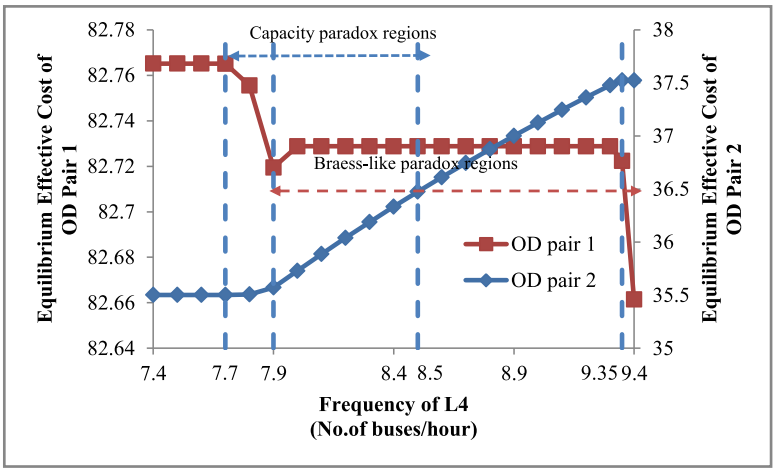


Fig. 5. Equilibrium effective travel costs of the two OD pairs under various frequencies of L4.

Table 3
Occurrence of the paradoxes.

	Before L4 is added	After L4 is added
Frequency of L4	–	9.0 buses/hour
Network residual capacity	360	0.0
Total effective travel cost	$44,932.8	$45,076.9

and the total effective travel costs are plotted in Fig. 4. The corresponding equilibrium effective travel costs of the two OD pairs and the flows on the three paths connecting OD pair 1 are plotted in Figs 5 and 6, respectively. The following discussions focus on the intuitive interpretation of the capacity paradox, and Appendix C provides the mathematical derivations of the condition for the existence of a more general parameter setting.

To clearly illustrate the results, five regions of the frequency of L4 are marked in Figs 4, 5, and 6: (1) less than 7.7 buses/hour (and greater than 0 buses/hour); (2) between 7.7 and 7.9 buses/hour; (3) between 7.9 and 8.5 buses/hour; (4) between 8.5 and 9.35 buses/hour; and (5) more than 9.35 buses/hour.

In the first region, the total effective travel cost and network residual capacity remain stable, indicating that increasing the frequency of line L4 does not affect the network performance in terms of the total effective travel cost or the network capacity. When the frequency of line L4 is less than 7.7 buses/hour, the route consisting of line L4 (i.e., path 3) has a higher effective travel cost than other routes. Therefore, none of the passengers travel on path 3. Fig. 6 illustrates this route choice behavior, showing that the flow on path 3 is zero and the flows on paths 1 and 2 do not change. Such a phenomenon implies that the line L4 buses would be vacant, and the operator would lose money by providing the L4 service at a low frequency.

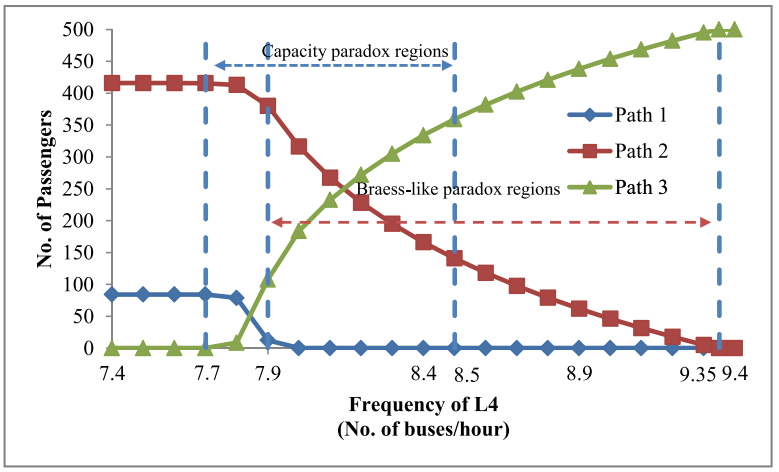


Fig. 6. Number of passengers on the paths connecting OD pair 1 under various frequencies of L4.

In the second region, the capacity paradox occurs; the total effective travel cost decreases as the frequency increases. The capacity paradox occurs because of the selfish route choice behavior of the passengers of OD pair 1. A number of passengers of OD pair 1 switch their routes from paths 1 and 2 to path 3 when the frequency of line L4 increases, because using path 3 can reduce the equilibrium effective travel cost of OD pair 1. However, the increment in the flow on path 3 increases the flows of all of the critical sections of the network simultaneously, as path 3 passes one of the critical sections, S3, and section S4, which competes with another critical section, S2. Consequently, the residual capacity of the network is reduced and the capacity paradox occurs. Meanwhile, the selfish behavior of the passengers of OD pair 1 generates additional congestion costs to the passengers of OD pair 2, as reflected by the increasing curve of the equilibrium effective travel cost of OD pair 2 in Fig. 5. Nevertheless, the increment in the equilibrium effective travel cost of OD pair 2 is mild compared to the reduction in that of OD pair 1, and the demand of OD pair 1 is higher than that of OD pair 2. Therefore, the reduction in the total effective travel cost of the system is driven by the reduction in the total effective travel cost of OD pair 1. This result implies that improving the network capacity can conflict with the aim of reducing the total effective travel cost of a network.

The third region is the region wherein the capacity and the Braess-like paradoxes occur simultaneously. The capacity paradox occurs because of the selfish behavior of passengers. With the increasing frequency of line L4, the effective travel cost on path 3 is reduced; thus, more passengers who were originally traveling on path 1 or 2 switch to path 3, as observed in Fig. 5. As a result, the flows on the critical sections (i.e., sections S3 and S2) keep increasing, resulting in a reduction in the network residual capacity. At the same time, the increased flow on path 3 induces higher congestion cost on section S3, and thus the equilibrium effective travel cost of OD pair 2, as shown in Fig. 5. Interestingly, the equilibrium effective travel cost of OD pair 1 in Fig. 5 shows various patterns. It first decreases when the frequency increases from the 7.9 to 8.0 buses/hour, then remains unchanged when the frequency is between 8.0 and 8.5 buses/hour. Such a trend can be explained as follows. A higher frequency can reduce waiting and congestion costs, resulting in a lower effective travel cost associated with path 3. In turn, a lower effective travel cost attracts more passengers from paths 1 and 2. However, more flow leads to a higher congestion cost. Therefore, the trend in the equilibrium effective travel cost of an OD pair depends on the overall effects of the reduction on the waiting and congestion costs due to the frequency increment, and the increment in the congestion cost due to the growth in the passenger flow. To be more specific, when the frequency increases from 7.9 to 8.0 buses/hour, the increment in the congestion cost caused by the growth in the passenger flow is larger than the reduction in waiting and congestion costs caused by the increment in the frequency; thus the equilibrium effective travel cost increases. When the frequency increases from 7.9 to 8.5 buses/hour, the reductions in the waiting and congestion costs caused by the increase in frequency counteract the increment in the congestion cost due to the growth in the passenger flow; thus, the equilibrium effective travel cost of OD pair 1 remains unchanged.

In the fourth region, the network residual capacity reduces to zero and the total effective travel cost keeps increasing. As the soft capacity constraint approach is adopted, passengers are still allowed to board fully occupied vehicles at the cost of bearing additional congestion cost. Therefore, in Fig. 6, the flow on path 3 keeps increasing and the equilibrium effective travel cost of OD pair 1 increases in Fig. 5. In contrast, the equilibrium effective travel cost of OD pair 1 remains stable until the end of this region, where the frequency equals 9.35 buses/hour, for reasons similar to those explained above. At the end of this region, when the frequency is 9.35 buses/hour, the total effective travel cost reaches its peak value, which is $45,113.5, because all of the passengers that originally traveled on path 1 or 2 have switched to path 3.

In the last region, where the frequency is greater than 9.35 buses/hour, the total effective travel cost begins to decrease. On the one hand, the increasing frequency pulls down the waiting cost. On the other hand, the congestion cost ceases increasing, as all of the passengers have changed their route to path 3, as observed in Fig. 6. The congestion cost decreases

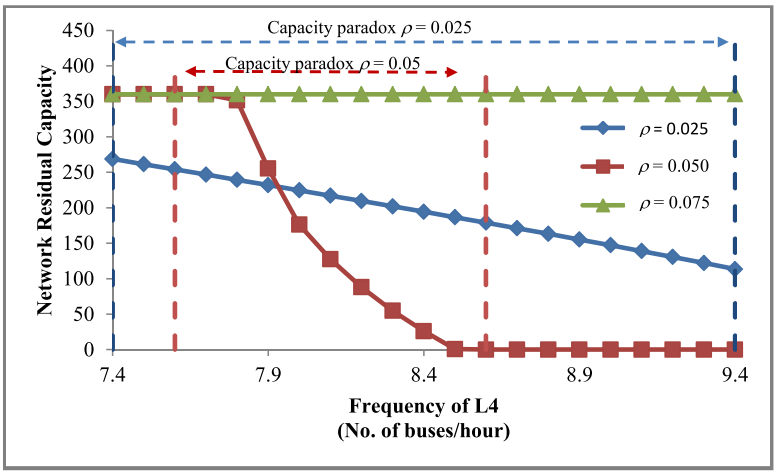


Fig. 7. Effect of the degree of risk aversion on the occurrence of the capacity paradox.

with the increasing frequency. Because the capacity of section S5 increases with frequency, the congestion cost obtained by the BPR type function is reduced. Therefore, the equilibrium travel cost of OD pair 1 reduces dramatically.

4.2.2. Implications

Fig. 4 shows three possible cases: (1) neither paradox occurs; (2) both paradoxes occur simultaneously; and (3) only one of the paradoxes occurs. Which case occurs depends on the initial and final frequencies. For example, if both frequencies fall in the second (fourth) region, then only the capacity (Braess-like) paradox occurs. However, if both fall in the third region, both paradoxes occur.

The above results imply that focusing on a single objective when improving the transit network (such as the total effective travel cost) may actually worsen the performance of another objective (such as network capacity). Hence, to improve transit networks, transit network design should simultaneously consider the total effective travel cost, network capacity, and route choice behavior, to avoid the occurrence of any of the two paradoxes. This calls for a need to solve a bilevel transit network design problem, which is left for future study.

4.2. Effect of the degree of risk aversion on the occurrence of the capacity paradox

Fig. 7 presents the changes in the network residual capacities when the degree of risk aversion parameter increases from 0.025 to 0.075. The network residual capacity decreases in the cases of $\rho = 0.025$ and $\rho = 0.05$. As the frequency of L4 increases, the waiting time cost is reduced, and the flow on path 3 increases. Hence, the capacity is reduced. The capacity is constant when $\rho = 0.075$, because path 3 still has a higher effective travel cost than path 2, the only path that carries flow, even if the waiting time cost is reduced. Therefore, when the value of ρ is high, the flow on path 3 is zero and the capacity paradox does not occur. Interestingly, it can be observed that the value of ρ also affects the shape of each of the curves depicting the changes in the network residual capacity. For example, when $\rho = 0.025$, the network residual capacity linearly reduces as the frequency increases, whereas it reduces nonlinearly when $\rho = 0.05$. Moreover, when $\rho = 0.05$, the network residual capacity reduces to zero, whereas it does not when $\rho = 0.025$. Furthermore, the value of ρ affects the width of the paradox region. For example, when $\rho = 0.05$, the paradox region starts at a frequency of 7.6 buses/hour and ends at 8.6 buses/hour, whereas the paradox region starts at a frequency of 7.4 buses/hour and continues to end of the tested region when $\rho = 0.025$. Essentially, the value of ρ determines the effective travel cost, the route choice, and network residual capacity and hence the occurrence of the capacity paradox (including the starting and ending frequencies). Therefore, it is important to accurately calibrate the degree of risk aversion parameter to determine whether the capacity paradox will occur. If it is unavoidable, managers should consider alternative methods to enhance system performance.

4.3. Effect of the perceived congestion parameter on the occurrence of the capacity paradox

The effect of the perceived congestion parameter, n, on the occurrence of the capacity paradox is demonstrated in Fig. 8, where the value of the perceived congestion parameter n is varied from 1 to 3. It can be seen that the network residual capacity is reduced when the frequency increases. The reduction in the (residual) network capacity is due to flow shift from paths 1 or 2 to path 3. Moreover, different trends in the reduction in the network capacity reveal different rates of flow shift. Fig. 8 indicates that when the perceived congestion cost is low, i.e., n is small, passengers change their routes more rapidly than in a heavily congested network, because path 3 is perceived as less congested than the other routes. Hence, more flow shifts to route 3 when n is small.

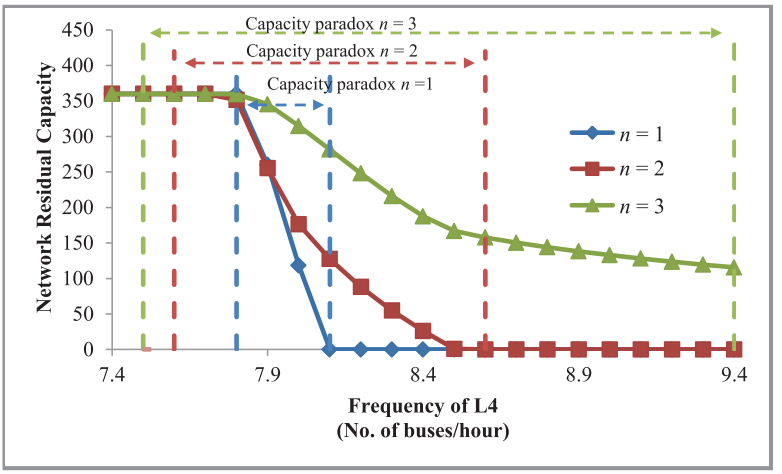


Fig. 8. Effect of the perceived congestion parameter on the occurrence of the capacity paradox.

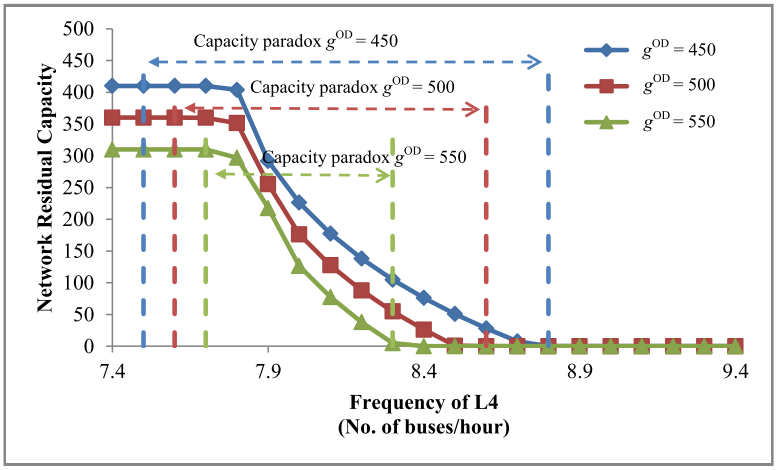


Fig. 9. Effect of demand on the occurrence of the capacity paradox.

As with the value of ρ, the value of the perceived congestion parameters should be properly calibrated to detect the existence of the capacity paradox, as these parameters also affect effective travel cost and hence route choice.

4.4. Effect of demand on the occurrence of the capacity paradox

Fig. 9 shows the effect of demand on the occurrence of the capacity paradox. In the figure, the demand of OD pair 1 varies from 450 to 550 passengers/hour. As expected, the demand level affects the initial residual capacity (i.e., the residual capacity when frequency equals 7.4 buses per hour). The higher the demand level, the lower the initial residual capacity. Furthermore, the lower the initial residual capacity, the smaller the frequency at which the network residual capacity reduces to zero. For example, when demand is 450 passengers/hour, the initial residual capacity is 410 passengers, and the residual capacity reduces to zero when the frequency is 8.8 buses/hour, whereas when the initial residual capacity is 310 passengers and the demand is 550 passengers, the residual capacity reduces to zero when the frequency is 8.4 buses/hour. Moreover, the range of frequency for the occurrence of the capacity paradox is wider for a lower demand. This is reasonable, as the initial residual network capacity is higher but the starting frequency is the same in all three cases. These results seem to indicate that the capacity paradox is more likely to occur when the demand is lower.

4.5. Comparison of the results with and without supply uncertainty

Table 4 demonstrates the effect of supply uncertainty on equilibrium flow and cost and the difference between the link-based and approach-based solutions. In this test, the value of ρ is set to 0.05. The frequency of the new line is set to 8.0 buses/hr. The table shows that only one section (i.e., section S1) leaving the origin carries flow when no supply uncertainty is considered, whereas two sections leaving the origin are used when supply uncertainty is considered. This implies

Table 4
Effect of supply uncertainty.

	With uncertainty		Without uncertainty	
	Solution	Equilibrium cost	Solution	Equilibrium cost
Link-based formulation	$v_{S1} = 0.0$	$u_{S1}^{OD} = 82.9$	$v_{S1} = 500.0$	$u_{S1}^{OD} = 53.7$
	$v_{S2} = 316.2$	$u_{S2}^{OD} = 82.7$	$v_{S2} = 0.0$	$u_{S2}^{OD} = 72.9$
	$v_{S3} = 183.8$	$u_{S3}^{BD} = 35.7$	$v_{S3} = 500.0$	$u_{S3}^{BD} = 24.8$
	$v_{S4} = 183.8$	$u_{S4}^{OD} = 82.7$	$v_{S4} = 0.0$	$u_{S4}^{OD} = 57.3$
	$v_{S5} = 183.8$	$u_{S5}^{AD} = 50.2$	$v_{S5} = 0.0$	$u_{S5}^{AD} = 36.3$
Approach-based formulation	$\alpha_{S1}^{d} = 0.00$	$u_{S1}^{OD} = 82.9$	$\alpha_{S1}^{d} = 1.00$	$u_{S1}^{OD} = 53.7$
	$\alpha_{S2}^{d} = 0.63$	$u_{S2}^{OD} = 82.7$	$\alpha_{S2}^{d} = 0.00$	$u_{S2}^{OD} = 72.9$
	$\alpha_{S3}^{d} = 1.00$	$u_{S3}^{BD} = 35.7$	$\alpha_{S3}^{d} = 1.00$	$u_{S3}^{BD} = 24.8$
	$\alpha_{S4}^{d} = 0.37$	$u_{S4}^{OD} = 82.7$	$\alpha_{S4}^{d} = 1.00$	$u_{S4}^{OD} = 57.3$
	$\alpha_{S5}^{d} = 1.00$	$u_{S5}^{AD} = 50.2$	$\alpha_{S5}^{d} = 1.00$	$u_{S5}^{AD} = 36.3$

that if uncertainty is not considered, the number of sections carrying flow could be underestimated and thus the flow on used bus routes could be overestimated. Nevertheless, the overestimation of the flow does not necessarily lead to a higher equilibrium cost, as the variance of the path travel cost is not captured in the equilibrium cost calculation.

A comparison of the solutions for the link-based and approach-based formulations shows that their equilibrium costs are equal, because the approach-based formulation is theoretically derived from the link-based formulation. However, the solutions of the two formulations are different. Specifically, a zero link flow may not result in a zero approach proportion, due to the definitional constraint of approach proportions (i.e., Eq. (32)). For example, under the column "Without Uncertainty," the link flow of section S4, v_{S4}, is zero in the link-based formulation, whereas in the approach-based formulation the proportion of section S4, α_{S4}^{d}, is 1.0.

4.6. Computational performance

The proposed solution method is tested using the Winnipeg network. The data for this network are obtained from the base scenario in Emme3, which consists of 1067 nodes, 3647 OD pairs, and 133 transit lines. As the original network is used for a multimodal assignment, some OD pairs are not connected by transit lines. For simplicity, these OD pairs are eliminated. The proposed algorithm is compared to the self-regulated averaging method (SAM) developed by Liu et al. (2009), which is used to solve the link-based problem. The SAM includes the method of successive averages (MSA) as a special case. Similar to the extragradient method, the SAM method uses two stepsize parameters for solution updates to increase the convergence speed. When both stepsizes are equal to 1, the SAM becomes equal to the MSA.

In the test, the following two gap measures are evaluated:

$$G_1 = \max\left[\delta_s^{t(s)d}\left(u_s^{t(s)d} - u^{t(s)d}\right), \forall s \in S, d \in D\right], \text{ and} \tag{38}$$

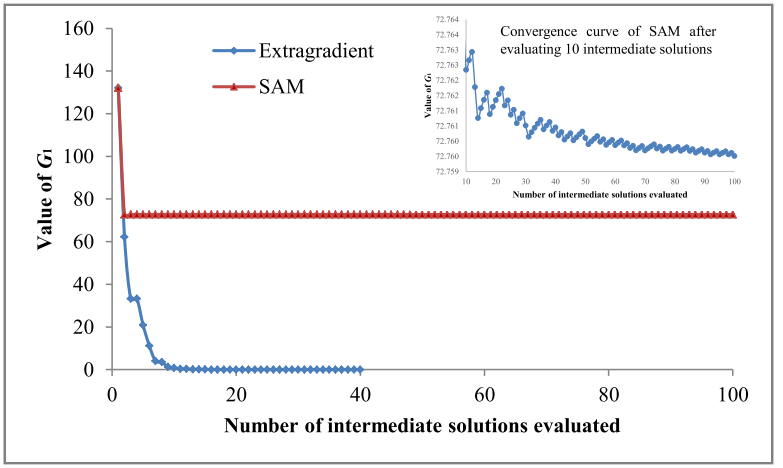
$$G_2 = \frac{|\mathbf{v}^l - \mathbf{v}^{l-1}|}{\mathbf{v}^{l-1}}, \tag{39}$$

where $\delta_s^{t(s)d}$equals 1 if section s carries flow (i.e., flow proportions greater than zero), and otherwise zero, and $\mathbf{v}^l$ is the l^{th} intermediate solution. G_1 is defined using the equilibrium conditions, whereas G_2 measures the difference in the link flow vectors between two consecutive iterations and is widely used in the MSA (e.g., Cantarella et al., 2015a,b).

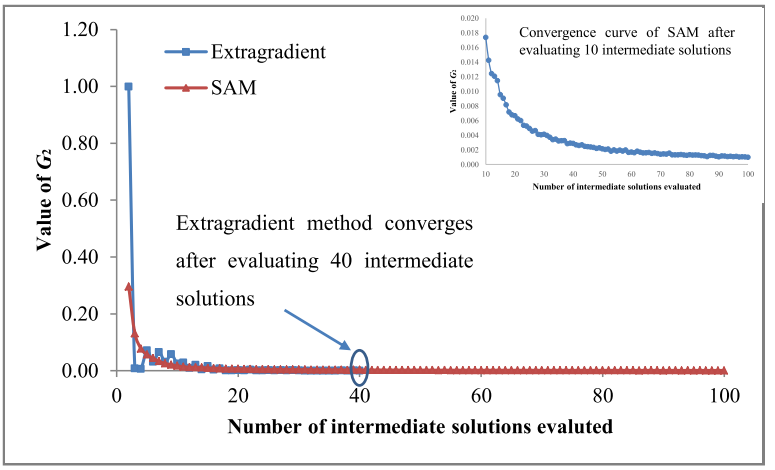
The convergence curves of the two measures are presented in Fig. 9. The x-axis plots the number of intermediate solutions evaluated and the y-axis represents the value of G_1 (Fig. 10a) or G_2 (Fig. 10b). Fig. 10 demonstrates that both measures converge faster with the extragradient method than with the SAM. More importantly, the SAM fails to converge using G_1, implying that the solution obtained from SAM might not satisfy the user equilibrium conditions. In addition, it is noted that the convergence curves of the extragradient method are not smooth, probably because the stepsizes can be too large in some iterations, leading to worse gap values.

5. Conclusions

This study develops a link-based variational inequality (VI) formulation for the reliability-based transit assignment problem that captures the stochasticity in the travel cost caused by supply uncertainty. An augmented route-section network representation is developed to internalize the covariance of the in-vehicle travel costs of the same line between different sections so that the resultant problem can be solved as if it was a link-based VI without supply uncertainty. This study also proves that in the link-based VI formulation, the mapping function is not strictly monotone. Thus, to solve the problem, the



(a) Convergence curve using measurement G_1



(b) Convergence curve using measurement G_2

Fig. 10. Comparison of convergence with the extragradient method and the SAM.

link-based formulation is first reformulated as an approach-based formulation in which the proportion of flow on each outgoing link from each node is a decision variable, and then an extragradient method, which only requires mild assumptions for convergence, is used. The result shows that the developed solution method can achieve a convergent solution, whereas the self-regulated averaging method, which includes the method of successive averages as a special case, may not. In addition, the result shows that without considering supply uncertainty, the number of route-sections carrying flow and the equilibrium cost could both be underestimated.

This study also introduces and investigates a capacity paradoxical phenomenon, in which adding a new line to a transit network or increasing the frequency of an existing transit line may reduce the network maximum throughput. The numerical results illustrate that the capacity paradox may or may not occur simultaneously with the Braess-like paradox. Moreover, both the degree of risk aversion of passengers and the demand level not only affect the occurrence of the capacity paradox but also determine the range of frequency of an existing line within which the capacity paradox could be observed.

This study opens up the following future research directions. 1) The BRP-type function implicitly assumes that all passengers have the same congestion cost, which may not be true given the different perceptions of sitting and standing passengers. Thus, one future direction is to explore and calibrate other cost functions, such as spline functions, that can capture this feature. Once a calibrated function that satisfies the pseudomonotonicity requirement is developed, it can be used in our proposed framework without conceptual difficulty. 2) This study only captures the first and second moments of travel time distribution. Empirical studies have pointed out that the transit travel time distribution can be highly skewed (Mazloumi et al., 2009; Xue et al., 2011). Thus, it is necessary to develop a transit assignment model that incorporates the third or even higher moments of travel time distribution, and to develop an efficient solution method to solve for solutions. 3) The demonstration of the capacity paradox calls for a bilevel multi-objective transit network design formulation to determine an optimal frequency to improve system performance in terms of the network maximum throughput and the total effective

travel cost. For example, the upper level could determine a Pareto optimal frequency that simultaneously minimizes the total effective travel cost and maximizes the network's reserved capacity, whereas the lower level is the approach-based transit assignment problem. 4) The model could also be extended to consider the stochastic user equilibrium.

Acknowledgments

This research was jointly supported by a grant from the University Research Committee of the University of Hong Kong (No. 201411159063), and a grant from the National Natural Science Foundation of China (No. 71271183). The authors are grateful to the two reviewers for their constructive comments.

Appendix A. Proof of the internalization of covariance using an augmented route-section network

Without loss of generality, let a path from node n_e to destination d be $p = (n_e \xrightarrow{s_e} n_{e+1} \xrightarrow{s_{e+1}} \cdots n_{e'} \xrightarrow{s_{e'}} n_{e'+1} \cdots \xrightarrow{s_{np-1}} n_{np} = d)$. $n_e, n_{e+1}, \ldots, n_{np}$ represent the nodes on path p. $s_e, s_{e+1} \ldots, s_{np-1}$ represent the sections on path p, where s_e and $s_{e'}$ indicate that these two sections contain the same line l.

In the link-based formulation, the random travel cost from nodes n_e to d via path p, $C_{s_e}^{n_e d}$, is defined by

$$C_{s_e}^{n_e d} = C_{s_e} + C^{n_{e+1} d}, \tag{A.1}$$

where C_{s_e} represents the random travel cost on section $s_e$and $C^{n_{e+1}d}$ is the random travel cost between node n_{e+1} and destination d. Eq. (A.1) means that the travel cost from node n_e to d via section s_e is the sum of the cost associated with section s_e and the travel cost from node n_{e+1}, which is the head node of section s_e, to destination d.

The effective travel cost is defined by

$$u_{s_e}^{n_e d} = E[C_{s_e}^{n_e d}] + \rho Var[C_{s_e}^{n_e d}], \tag{A.2}$$

where the expectation and variance are respectively derived as

$$E[C_{s_e}^{n_e d}] = E[C_{s_e} + C^{n_{e+1} d}] = E[C_{s_e}] + E[C^{n_{e+1} d}] \tag{A.3}$$

and

$$Var[C_{s_e}^{n_e d}] = Var[C_{s_e} + C^{n_{e+1} d}] = Var[C_{s_e}] + Var[C^{n_{e+1} d}] + 2Cov[C_{s_e}, C^{n_{e+1} d}]. \tag{A.4}$$

By substituting Eqs (A.3) and (A.4) into Eq. (A.2), it can be obtained that

$$\begin{aligned} u_{s_e}^{n_e d} &= E[C_{s_e}] + E[C^{n_{e+1} d}] + \rho \begin{pmatrix} Var[C_{s_e}] + Var[C^{n_{e+1} d}] \\ +2Cov[C_{s_e}, C^{n_{e+1} d}] \end{pmatrix} \\ &= (E[C_{s_e}] + \rho Var[C_{s_e}]) + (E[C^{n_{e+1} d}] + \rho Var[C^{n_{e+1} d}]) \\ &\quad +2\rho Cov[C_{s_e}, C^{n_{e+1} d}]. \end{aligned} \tag{A.5}$$

The first round bracket term $E[C_{s_e}] + \rho Var[C_{s_e}]$ in the last equality is the effective cost associated with section s_e, the second round bracket term is the effective cost between nodes n_{e+1} and d, or $u^{n_{e+1} d}$, and the third term is the covariance between the travel cost associated with section s_e and the travel cost from node n_{e+1} to d.

The following shows how to address the covariance term using the augmented route-section network under the assumption that passengers prefer a direct service and do not transfer within the same line.

First, the variance $Var[C_{s_e}^{n_e d}]$ in Eq. (A.5) can be rewritten in the following format:

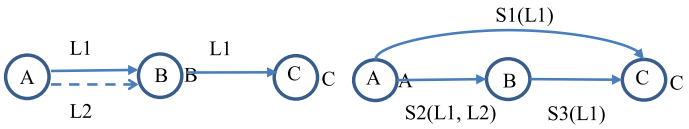
$$Var\left[C_{s_e}^{n_e d}\right] = Var\left[\sum_{s'} \Delta_{s'}^{n_e d, p} C_{s'}\right], \tag{A.6}$$

where $\Delta_{s'}^{n_e d, p} = 1$ if section s' is on path p connecting nodes n_e and d; otherwise $\Delta_{s'}^{n_e d, p} = 0$. Then, the variance can be expressed as

$$Var\left[C_{s_e}^{n_e d}\right] = Var\left[\sum_{s'} \Delta_{s'}^{n_e d, p} C_{s'}\right] = \sum_{s'} \Delta_{s'}^{n_e d, p} Var[C_{s'}] + \sum_{s'} \sum_{s'' \neq s'} \Delta_{s'}^{n_e d, p} \Delta_{s''}^{n_e d, p} Cov[C_{s'}, C_{s''}]. \tag{A.7}$$

Second, based on the assumptions that only the covariance of the same line between different sections are non-zero and that all of the other waiting times, in-vehicle travel times, and perceived congestion costs are independent, Eq. (A.7) can be reduced to

$$Var\left[C_{s_e}^{n_e d}\right] = \sum_{s'} \Delta_{s'}^{n_e d, p} Var[C_{s'}] + Cov[T_{s_e}, T_{s_{e'}}]$$



(a) Transit network representation using transit lines (b) Transit network representation using route sections

Fig. A. The small network for the proof of Proposition 1.

$$= \sum_{s'} \Delta_{s'}^{n_e d,p} Var[C_{s'}] + Cov\left[\sum_{l'} w_{s_e}{}^{l'} T_{s_e}{}^{l'}, \sum_{l'} w_{s_{e'}}{}^{l'} T_{s_{e'}}{}^{l'}\right]$$
$$= \sum_{s'} \Delta_{s'}^{n_e d,p} Var[C_{s'}] + w_{s_e}{}^{l} w_{s_{e'}} Cov[T_{s_e}, T_{s_{e'}}] \tag{A.8}$$

$Cov[T_{s_e}, T_{s_{e'}}]$ is incorporated, as sections s_e and $s_{e'}$ contain the same line l. $Cov[\sum_{l'} w_{s_e}^{l'} T_{s_e}^{l'}, \sum_{l'} w_{s_{e'}}^{l'} T_{s_{e'}}^{l'}]$ is obtained based on Eq. (2).

To analyze Eq. (A.8), the following two scenarios are considered.

Scenario 1: If $w_{s_e}^{l} w_{s_{e'}}^{l} = 0$, the term $w_{s_e}^{l} w_{s_{e'}}^{l} Cov[T_{s_e}^{l}, T_{s_{e'}}^{l}]$ becomes zero and is eliminated. In fact, such a scenario implies that line l is not an attractive line on section s_e or $s_{e'}$ or both s_e and $s_{e'}$.

Scenario 2: If $w_{s_e}^{l} w_{s_{e'}}^{l} \neq 0$, the term $w_{s_e}^{l} w_{s_{e'}}^{l} Cov[T_{s_e}^{l}, T_{s_{e'}}^{l}]$ can be addressed by the augmented route-section network and under the assumption that passengers prefer a direct service and do not transfer within the same line. The explanation is as follows.

In the augmented route-section network, there must exist a dummy section to penalize a transfer within the same line. Hence, path p has a large cost and is eliminated by passengers if there is another path. However, in the augmented route-section network, there must exist a section connecting nodes n_e and $n_{e'+1}$ directly, as line l traverses both nodes. Hence, a new path p' that avoids transfers and connects n_e and $n_{e'+1}$ must exist. Therefore, the variance of the new path p' is analyzed instead of that of path p. Path p' is represented as $p' = (n_e \xrightarrow{s'_e} n_{e'+1} \cdots \xrightarrow{s_{n^{p-1}}} n^p = d)$, where s'_e is the section that connects nodes n_e and $n_{e'+1}$ directly. The variance of the cost associated with path p' is given by

$$Var\left[C_{s'_e}^{n_e d}\right] = Var\left[\sum_{s'} \Delta_{s'}^{n_e d,p'} C_{s'}\right] = \sum_{s'} \Delta_{s'}^{n_e d,p'} Var[C_{s'}]. \tag{A.9}$$

Note that there is no covariance term in (A.9) because only section s'_e contains line l, and its in-vehicle travel time is assumed to be independent of that of other lines. Hence, $w_{s_e}^{l} w_{s_{e'}}^{l} Cov[T_{s_e}^{l}, T_{s_{e'}}^{l}]$ does not need to be considered, as the path including this term carries no flow at optimality. In fact, the covariance of line l between sections s_e and $s_{e'}$ has been internalized into the variance of section s'_e, which can be computed in advance.

To sum up, in the augmented route-section network under the assumption that passengers prefer a direct service and do not make a transfer within the same line, if there is path p connecting nodes n_e and d, such as $p = (n_e \xrightarrow{s_e} n_{e+1} \xrightarrow{s_{e+1}} \cdots n_{e'} \xrightarrow{s_{e'}} n_{e'+1} \cdots \xrightarrow{s_{n^{p-1}}} n_{n^p} = d)$, a new route $p' = (n_e \xrightarrow{s'_e} n_{e'+1} \cdots \xrightarrow{s_{n^{p-1}}} n^p = d)$ is adopted to replace route p. As a result, any route can only include a line at most once. Hence, the expectation and variance can be respectively computed by

$$E\left[C_{s'_e}^{n_e d}\right] = E\left[C_{s'_e} + C^{n_{e'+1} d}\right] = E[C_{s'_e}] + E[C^{n_{e'+1} d}] \quad \text{and} \tag{A.10}$$

$$Var\left[C_{s'_e}^{n_e d}\right] = Var\left[C_{s'_e} + C^{n_{e'+1} d}\right] = Var[C_{s'_e}] + Var\left[C^{n_{e'+1} d}\right]. \tag{A.11}$$

The effective travel cost from node n_e to d via section s'_e is obtained by

$$u_{s'_e}^{n_e d} = (E[C_{s'_e}] + \rho Var[C_{s'_e}]) + u^{n_{e'+1} d}. \tag{A.12}$$

Appendix B. Proof of the non-monotonicity of the mapping function and the existence of multiple solutions

This appendix gives the proofs of Propositions 1 and 2.

Proposition 1. *The mapping function in the link-based VI formulation is not strictly monotone.*

Proof: In this appendix, a small network is created in Fig. A to demonstrate that the mapping function in the link-based VI formulation is not strictly monotone.

The network contains two lines: L1 and L2. Fig. A(a) is network representation using transit lines and Fig. A(b) is the route-section network representation. Consider OD pair AC, there are two sections emanating from origin A heading to destination C. One is via section S1 and the other is via section S2. Denote $\mathbf{v} = \begin{pmatrix} v_1^C \\ v_2^C \end{pmatrix}$ as the vector of link flows and $\mathbf{u}$ as the vector of the corresponding mapping functions.

If the mapping function is strictly monotone, then the following condition should be satisfied:

$$\left(\mathbf{u} - \mathbf{u}'\right)^T \left(\mathbf{v} - \mathbf{v}'\right) > 0, \forall \mathbf{v}, \mathbf{v}' \in \Omega_v, \tag{B.1}$$

where $\Omega_v = (v_1^C, v_2^C | v_1^C, v_2^C \geq 0, \text{and } v_1^C + v_2^C = g^{AC})$ is the solution space and g^{AC} is the demand of OD pair AC.

To demonstrate that the mapping function is not strictly monotone, we will show that the following equation holds under a certain parameter setting:

$$\left(\mathbf{u} - \mathbf{u}'\right)^T \left(\mathbf{v} - \mathbf{v}'\right) = 0, \forall \mathbf{v}, \mathbf{v}' \in \Omega_v. \tag{B.2}$$

Our proof contains the following steps.

Step 1. Derive the mapping functions

The mapping function can be defined as

$$\mathbf{u} = \begin{pmatrix} u_1^C \\ u_2^C \end{pmatrix}, \tag{B.3}$$

where u_1^C and u_2^C are obtained by

$$u_1^C = E[C_1] + \rho Var[C_1] \quad \text{and} \tag{B.4}$$

$$u_2^C = E[C_2 + C_3] + \rho Var[C_2 + C_3]. \tag{B.5}$$

Based upon Eqs. (1) to (9), the cost components are expressed as

$$E[C_1] = E[T_1 + W_1 + \Phi_1 + T_{1,0}] = E[T_1] + E[W_1] + E[T_{1,0}] + E[\Phi_1] \tag{B.6}$$

$$Var[C_1] = Var[T_1 + W_1 + \Phi_1 + T_{1,0}] = Var[T_1] + Var[W_1] + Var[T_{1,0}] + Var[\Phi_1] \tag{B.7}$$

$$\begin{aligned} E[C_2 + C_3] = {} & E[T_2] + E[W_2] + E[T_{2,0}] + E[\Phi_2] \\ & + E[T_3] + E[W_3] + E[T_{3,0}] + E[\Phi_2], \quad \text{and} \end{aligned} \tag{B.8}$$

$$\begin{aligned} Var[C_2 + C_3] = {} & Var[T_2] + Var[W_2] + Var[T_{2,0}] + Var[\Phi_2] \\ & + Var[T_3] + Var[W_3] + Var[T_{3,0}] + Var[\Phi_3]. \end{aligned} \tag{B.9}$$

As only the expectation and variance of the congestion cost are flow dependent, we further derive the expectation and variance of congestion cost as

$$E[\Phi_1] = \mu_\Phi b_{1,1} n! \left(\frac{b_{1,2} v_1^C + b_{1,3} w_2^1 v_2^C}{f_1 \gamma_k k} \right)^n, \tag{B.10}$$

$$Var[\Phi_1] = \mu_\Phi^2 \left(b_{1,1}\right)^2 \left((2n)! - (n!)^2\right) \left(\frac{b_{1,2} v_1^C + b_{1,3} w_2^1 v_2^C}{f_1 \gamma_k k} \right)^{2n}, \tag{B.11}$$

$$\begin{aligned} E[\Phi_2 + \Phi_3] &= E[\Phi_2] + E[\Phi_3] \\ &= \mu_\Phi \left(b_{2,1} n! \left(\frac{b_{2,2} v_2^C + b_{2,3} v_1^C}{f_1 \gamma_k k} \right)^n + b_{3,1} n! \left(\frac{b_{3,2} v_3^C + b_{3,3} v_1^C}{f_1 \gamma_k k} \right)^n \right), \quad \text{and} \end{aligned} \tag{B.12}$$

$$Var[\Phi_2+\Phi_3] = Var[\Phi_2]+Var[\Phi_3]$$
$$= \mu_\Phi^2\left(b_{2,1}\right)^2\left((2n)!-(n!)^2\right)\left(\frac{b_{2,2}v_2^C+b_{2,3}v_1^C}{f_2\gamma_k k}\right)^{2n}$$
$$+\mu_\Phi^2\left(b_{3,1}\right)^2\left((2n)!-(n!)^2\right)\left(\frac{b_{3,2}v_3^C+b_{3,3}v_1^C}{f_3\gamma_k k}\right)^{2n}, \tag{B.13}$$

where v_3^C in Eq. (B.13) is the flow on Section 3 going to destination C and it equals v_2^C according to the flow conversation. $w_2{}^1$ is the relative frequency of line 1 on Section 2.

Step 2. Define a specific parameter setting

For the small example, the parameters are specified as $\mu_\Phi = 1$, $b_{2,2} = b_{2,3} = b_{3,2} = b_{3,3} = b_{1,2} = 1$, $f^1 = f^2 = 4$ buses/hour, and $b_{1,3} = 2$. Accordingly, the means and variances in Eqs. (B.10)–(B.13) reduce to

$$E[\Phi_1] = b_{1,1}n!\left(\frac{v_1^C+v_2^C}{f_1\gamma_k k}\right)^n, \tag{B.14}$$

$$Var[\Phi_1] = \left(b_{1,1}\right)^2\left((2n)!-(n!)^2\right)\left(\frac{v_1^C+v_2^C}{f_1\gamma_k k}\right)^{2n}, \tag{B.15}$$

$$E[\Phi_2+\Phi_3] = b_{2,1}n!\left(\frac{v_2^C+v_1^C}{f_2\gamma_k k}\right)^n + b_{3,1}n!\left(\frac{v_3^C+v_1^C}{f_3\gamma_k k}\right)^n, \quad \text{and} \tag{B.16}$$

$$Var[\Phi_2+\Phi_3] = \left(b_{2,1}\right)^2\left((2n)!-(n!)^2\right)\left(\frac{v_2^C+v_1^C}{f_2\gamma_k k}\right)^{2n}$$
$$+\left(b_{3,1}\right)^2\left((2n)!-(n!)^2\right)\left(\frac{v_3^C+v_1^C}{f_3\gamma_k k}\right)^{2n}. \tag{B.17}$$

In addition, based on the flow conservation conditions, we have

$$v_1^C+v_2^C = g^{AC} \quad \text{and} \tag{B.18}$$

$$v_3^C+v_1^C = g^{AC}. \tag{B.19}$$

By substituting Eqs. (B.18) and (B.19) into Eqs. (B.14) to (B.17), it is obtained that the mean and variance of the congestion cost become functions of demand g^{AC} and independent of $\mathbf{v} \in \Omega_v$. Therefore, u_1^C and u_2^C become functions of demand g^{AC} and independent of $\mathbf{v} \in \Omega_v$. Hence, when the demand g^{AC} is fixed, for any $\mathbf{v} \in \Omega_v$, we can represent the mapping function as

$$\mathbf{u}(\mathbf{v}) = \phi\left(g^{AC}\right) = \mathbf{K}, \forall \mathbf{v} \in \Omega_v, \tag{B.21}$$

where $\boldsymbol{\phi}(g^{AC})$ is a function of demand and $\mathbf{K}$ represents a vector of constants.

Therefore, we have

$$\left(\mathbf{u}(\mathbf{v})-\mathbf{u}(\mathbf{v}')\right)^T\left(\mathbf{v}-\mathbf{v}'\right) = (\mathbf{K}-\mathbf{K})^T\left(\mathbf{v}-\mathbf{v}'\right) = 0, \forall \mathbf{v}, \mathbf{v}' \in \Omega_v. \tag{B.22}$$

This completes the proof. □

Proposition 2. *Multiple solutions exist to the link-based VI formulation.*

Proof: Consider the same network and parameter setting as in the proof of Proposition 1. As the mapping function only depends on the demand level, any flow pattern satisfying the flow conservation constraint and nonnegativity constraint results in the same value for the mapping function. Therefore, multiple solutions exist to the problem.

This completes the proof. □

Appendix C. Conditions for the occurrence of the capacity paradox

This appendix derives the conditions for the occurrence of the capacity paradox by comparing the following two scenarios. The first is the scenario before the new line is provided (i.e., the frequency of line L4 is zero), and the second is the scenario after the new line is added (i.e., the frequency of line L4 is positive).

Scenario 1: Before adding the new line

In this scenario, there are two used routes, paths 1 and 2, connecting nodes O and D. Denote f_1^{OD} and f_2^{OD} as the flow on paths 1 and 2, respectively. The following flow conservation and nonnegativity constraints hold:

$$f_1^{OD} + f_2^{OD} = g^{OD}. \tag{C.1}$$

For path 2, the flow passes only one section, S2; thus, S2 is the critical section. For path 1, w.l.o.g., let section S3 be the critical section by setting $E[K_{S1}] > E[K_{S3}]$, where $E[K_{S3}]$ and $E[K_{S1}]$ represent the expected capacity of sections S1 and S3, respectively. The network residual capacity in scenario 1, $E[K_{residual}]_1$, is calculated by

$$E\left[K_{residual}\right]_1 = E[K_{S3}] - v_{S3} + E[K_{S2}] - v_{S2}, \tag{C.2}$$

where v_{S3} and v_{S2} are, respectively, the flows on sections S3 and S2 in scenario 1 and are determined by

$$v_{S3} = f_1^{OD} \quad \text{and} \tag{C.3}$$

$$v_{S2} = f_2^{OD}. \tag{C.4}$$

By substituting Eqs. (C.3) and (C.4) into Eq. (C.2), the network residual capacity of scenario 1 can be represented as

$$E[K_{residual}]_1 = E[K_{S3}] + E[K_{S2}] - f_1{}^{OD} - f_2{}^{OD} = E[K_{S3}] + E[K_{S2}] - g^{OD}. \tag{C.5}$$

The above equation implies that the network residual capacity is determined by the expected capacities of the two critical sections, $E[K_{S3}]$ and $E[K_{S2}]$, and demand g^{OD}.

Scenario 2: After adding the new line

Once the frequency of line L4 is greater than zero, nodes A and B are connected, and route 3 is available. Denote f_3^{OD} as the flow on route 3. The flow conservation and nonnegativity constraints can be revised as

$$f_1^{OD} + f_2^{OD} + f_3^{OD} = g^{OD} \tag{C.6}$$

Furthermore, the effective section flows on sections S2 and S3 in scenario 2 are also obtained by

$$v_{S3} = f_1^{OD} + f_3^{OD} \quad \text{and} \tag{C.7}$$

$$v_{S2} = f_2^{OD} + f_3^{OD}. \tag{C.8}$$

Eq. (C.7) means that section S3 contains the flows on paths 1 and 3. Eq. (C.8) states that the effective flow on Section 2 contains the flows on paths 2 and 3.

By substituting Eqs (C.7) and (C.8) into Eq. (C.2), the network residual capacity in scenario 2 can be expressed as

$$\begin{aligned} E[K_{residual}]_2 &= E[K_{S3}] - f_1^{OD} - f_3^{OD} + E[K_{S2}] - f_2^{OD} - f_3^{OD} \\ &= E[K_{S3}] + E[K_{S2}] - g^{OD} - f_3^{OD}. \end{aligned} \tag{C.9}$$

Comparing Eq. (C.9) with Eq. (C.5), it is observed that the difference between $E[K_{residual}]_1$ and $E[K_{residual}]_2$ is f_3^{OD}. It is concluded that if $f_3^{OD} > 0$, $E[K_{residual}]_2 < E[K_{residual}]_1$, indicating the occurrence of the capacity paradox. Furthermore, it is also concluded that if the network residual capacity is reduced, then the reduction in the residual capacity must equal f_3^{OD}.

References

Anderson, M.K., 2013. Behavioural Models for Route Choice of Passengers in Multimodal Public Transport Networks. Dissertation, Technical University of Denmark.

Anderson, M.K., Nielsen, O.A., Prato, C.G., 2014. Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. EURO J. Transp. Logist. doi:10.1007/s13676-014-0063-3.

Bar-Gera, H., 2002. Origin-based algorithm for the traffic assignment problem. Transp. Sci. 36 (4), 398–417.

Chriqui, C., 1974. Réseaux de Transports en Commun: Les Problèmes de Cheminement et d'Accès. Département d'Informatique, Université de Montréal.

Blum, J.R., 1954. Multidimensional stochastic approximation methods. Ann. Math. Stat. 25 (4), 737–744.

Cantarella, G.E., Cartenì, A., de Luca, S., 2015a. Stochastic equilibrium assignment with variable demand: theoretical and implementation issues. Eur. J. Oper. Res. 241 (2), 330–347.

Cantarella, G.E., de Luca, S., Di Gangi, M., Di Pace, R., 2015b. Approaches for solving the stochastic equilibrium assignment with variable demand: internal vs. external solution algorithms. Optim. Methods Softw. 30 (2), 338–364.

Carrel, A., Halvorsen, A., Walker, J., 2013. Passengers' perception of and behavioral adaptation to unreliability in public transportation. Transp. Res. Rec. 2351, 153–162.

Casello, J., Nour, A., Hellinga, B., 2009. Quantifying impacts of transit reliability on user costs. Transp. Res. Rec. 2112, 136–141.

Cepeda, M., Cominetti, R., Florian, M., 2006. A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. Transp. Res. Part B 40 (6), 437–459.

Chen, X., Yu, L., Zhang, Y., Guo, J., 2009. Analyzing urban bus service reliability at the stop, route, and network levels. Transp. Res. Part A 43 (8), 722–734.

Chen, Y.J., Li, Z.C, Lam, W.H.K., 2015. Modeling transit technology selection in a linear transportation corridor. J. Adv. Transp. 49 (1), 48–72.

Chriqui, C., Robillard, P., 1975. Common bus lines. Transp. Sci. 9 (2), 115–121.

Cominetti, R., Correa, J., 2001. Common-lines and passenger assignment in congested transit networks. Transp. Sci. 35 (3), 250–267.

Cortés, C.E., Jara-Moroni, P., Moreno, E., Pineda, C., 2013. Stochastic transit equilibrium. Transp. Res. Part B 51, 29–44.

de Cea, J., Fernández, E., 1993. Transit assignment for congested public transport systems: an equilibrium model. Transp. Sci. 27 (2), 133–147.

Diab, E., El-Geneidy, A., 2013. Variation in bus transit service: understanding the impacts of various improvement strategies on transit service reliability. Public Transp. 4 (3), 209–231.

Dial, R.B., 1967. Transit pathfinder algorithms. Highway Res. Rec. 205, 67–85.
Fearnside, K., Draper, D.P., 1971. Public transport assignment-a new approach. Traffic Eng. Control 12, 298–299.
Florian, M., 1998. Deterministic time table transit assignment. Preprints of PTRC Seminar on National Models, Stockholm.
Florian, M., 2004. Finding shortest time-dependent paths in schedule-based transit networks: a label setting algorithm. In: Wilson, N.H.M., Nuzzolo, A. (Eds.), Schedule-based Dynamic Transit Modeling: Theory and Applications. Springer, US, pp. 43–52.
Fosgerau, M., Engelson, L., 2011. The value of travel time variance. Transp. Res. Part B 45 (1), 1–8.
Frumin, M., Zhao, J., 2012. Analyzing passenger incidence behavior in heterogeneous transit services using smartcard data and schedule-based assignment. Transp. Res. Rec. 2274, 52–60.
Fu, X., Lam, W.H.K, Chen, B.Y., 2014. A reliability-based traffic assignment model for multi-modal transport network under demand uncertainty. J. Adv. Transp. 48 (1), 66–85.
Habib, K.M.N., Kattan, L., Islam, M.T., 2011. Model of personal attitudes towards transit service quality. J. Adv. Transp. 45 (4), 271–285.
Hamdouch, Y., Lawphongpanich, S., 2008. Schedule-based transit assignment model with travel strategies and capacity constraints. Transp. Res. Part B 42 (7), 663–684.
Hamdouch, Y., Szeto, W.Y., Jiang, Y., 2014. A new schedule-based transit assignment model with travel strategies and supply uncertainties. Transp. Res. Part B 67, 35–67.
Hoogendoorn-Lanser S., 2005. Modelling Travel Behaviour for Multi-Modal Transport Networks. TRAIL Thesis Series T2005/4, TRAIL, The Netherlands
Hoogendoorn-Lanser, S., van Nes, R., Bovy, P., 2005. Path size modeling in multimodal route choice analysis. Transp. Res. Rec. 1921, 27–34.
Jackson, W.B., Jucker, J.V., 1982. An empirical study of travel time variability and travel choice behavior. Transp. Sci. 16 (4), 460–475.
Korpelevich, G., 1976. The extragradient method for finding saddle points and other problems. Matecon 12, 747–756.
Kurauchi, F., Bell, M., Schmöcker, J.D., 2003. Capacity constrained transit assignment with common lines. J. Math. Modell. Algorithms 2, 309–327.
Johnson, G.G., 1972. Fixed points by mean value iterations. Proc. Am. Math. Soc. 34 (1), 193–194.
Last, A., Leak, S.E., 1976. Transept: a bus model. Traffic Eng. Control 17, 14–20.
Lam, W.H.K., Gao, Z.Y., Chan, K.S., Yang, H., 1999. A stochastic user equilibrium assignment model for congested transit networks. Transp. Res. Part B 33 (5), 351–368.
Lam, W.H.K., Shao, H., Sumalee, A., 2008. Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. Transp. Res. Part B 42 (10), 890–910.
Lam, W.H.K., Zhou, J., Sheng, Z.H., 2002. A capacity restraint transit assignment with elastic line frequency. Transp. Res. Part B 36 (10), 919–938.
Lampkin, W., Saalmans, P.D., 1967. The design of routes, service frequencies, and schedules for a municipal bus undertaking: a case study. Oper. Res. Q. 18, 375–397.
Le Clercq, F., 1972. A public transport assignment model. Traffic Eng. Control 13, 91–96.
Leurent, F., 2012. On seat capacity in traffic assignment to a transit network. J. Adv. Transp. 46 (2), 112–138.
Li, Z.C., Lam, W.H.K., Sumalee, A., 2008. Modeling impacts of transit operator fleet size under various market regimes with uncertainty in network. Transp. Res. Rec. 2063, 18–27.
Li, Z.C., Lam, W.H.K., Wong, S.C., 2009a. Optimization of a bus and rail transit system with feeder bus services under different market regimes. Transp. Traffic Theory 2009, 495–516.
Li, Z.C., Lam, W.H.K., Wong, S.C., 2009b. The optimal transit fare structure under different market regimes with uncertainty in the network. Netw. Spatial Econ. 9 (2), 191–216.
Liu, H.X., He, X.Z., He, B.S., 2009. Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem. Netw. Spatial Econ. 9 (4), 485–503.
Liu, Z., Meng, Q., 2014. Modelling transit-based park-and-ride services on a multimodal network with congestion pricing schemes. Int. J. Syst. Sci. 45 (5), 994–1006.
Lo, H.K., Luo, X.W., Siu, B.W.Y., 2006. Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. Transp. Res. Part B 40 (9), 792–806.
Lo, H.K., Wan, Q., Yip, C.W., 2002. Multimodal transit services with heterogeneous travelers. Transp. Res. Rec. 1799, 26–34.
Lo, H.K., Yip, C.W., Wan, K.H., 2003. Modeling transfer and non-linear fare structure in multi-modal network. Transp. Res. Part B 37 (2), 149–170.
Lo, H.K., Yip, C.W., Wan, Q.K., 2004. Modeling competitive multi-modal transit services: a nested logit approach. Transp. Res. Part C 12 (3), 251–272.
Lozano, A., Storchi, G., 2001. Shortest viable path algorithm in multimodal networks. Transp. Res. Part A 35 (3), 225–241.
Mabit, S.L., Nielsen, O.A., 2006. The effect of correlated value of travel time savings in public transport assignment. In: Proceedings of the European Transport Conference, 18 September 2006. Strasbourg, France.
Mazloumi, E., Currie, G., Rose, G., 2009. Using GPS data to gain insight into public transport travel time variability. J. Transp. Eng. 136 (7), 623–631.
Meng, Q., Qu, X., 2013. Bus dwell time estimation at bus bays: a probabilistic approach. Transp. Res. Part C 36, 61–71.
Nagurney, A., 1993. Network Economics: a Variational Inequality Approach. Kluwer Academic Publishers, Norwell, MA, USA.
Nguyen, S., Pallottino, S., 1988. Equilibrium traffic assignment for large scale transit networks. Eur. J. Oper. Res. 37 (2), 176–186.
Nguyen, S., Pallottino, S., Malucelli, F., 2001. A modeling framework for passenger assignment on a transport network with timetables. Transp. Sci. 35 (3), 238–249.
Nielsen, O.A, Frederiksen, R., 2006. Optimisation of timetable-based, stochastic transit assignment models based on MSA. Ann. Oper. Res. 144 (1), 263–285.
Nielsen, O.A., 2000. A stochastic transit assignment model considering differences in passengers utility functions. Transp. Res. Part B 34 (5), 377–402.
Nielsen, O.A., 2004. A large scale stochastic multi-class schedule-based transit model with random coefficients. In: Wilson, N.H.M., Nuzzolo, A. (Eds.), Schedule-Based Dynamic Transit Modeling: Theory and Applications. Springer, US, pp. 53–77.
Nielsen, O.A., Daly, A., Frederiksen, R.D., 2002. A stochastic route choice model for car travellers in the Copenhagen region. Netw. Spatial Econ. 2 (4), 327–346.
Nuzzolo, A., Crisalli, U., Rosati, L., 2012. A schedule-based assignment model with explicit capacity constraints for congested transit networks. Transp. Res. Part C 20 (1), 16–33.
Nuzzolo, A., Russo, F., Crisalli, U., 2001. A doubly dynamic schedule-based assignment model for transit networks. Transp. Sci. 35 (3), 268–285.
Rasmussen, T.K., Watling, D.P., Prato, C.G., Nielsen, O.A., 2015. Stochastic user equilibrium with equilibrated choice sets: part II–Solving the restricted SUE for the logit family. Transp. Res. Part B 77, 146–165.
Ren, H., Gao, Z., Lam, W.H.K., Long, J., 2009. Assessing the benefits of integrated en-route transit information systems and time-varying transit pricing systems in a congested transit network. Transp. Plann. Technol. 32 (3), 215–237.
Schmöcker, J.D., Fonzone, A., Shimamoto, H., Kurauchi, F., Bell, M.G.H., 2011. Frequency-based transit assignment considering seat capacities. Transp. Res. Part B 45 (2), 392–408.
Senna, L.A., 1994. The influence of travel time variability on the value of time. Transportation 21 (2), 203–228.
Shao, H., Lam, W.H.K., Tam, M.L., Yuan, X.M., 2008. Modelling rain effects on risk-taking behaviours of multi-user classes in road networks with uncertainty. J. Adv. Transp. 42 (3), 265–290.
Siu, B.W.Y., Lo, H.K., 2008. Doubly uncertain transportation network: degradable capacity and stochastic demand. Eur. J. Oper. Res. 191 (1), 166–181.
Spiess, H., 1984. Contributions à la Théorie et aux Outils de Planification de Réseaux de Transport Urbain. Ph.D. thesis, Département d'Informatique et de Recherche Opérationnelle, Centre de recherche sur les transports, Université de Montréal.
Spiess, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. Transp. Res. Part B 23 (2), 83–102.
Sumalee, A., Tan, Z., Lam, W.H.K., 2009. Dynamic stochastic transit assignment with explicit seat allocation model. Transp. Res. Part B 43 (8), 895–912.

Sumalee, A., Uchida, K., Lam, W.H.K., 2011. Stochastic multi-modal transport network under demand uncertainties and adverse weather condition. Transp. Res. Part C 19, 338–350.
Sun, L., Meng, Q., Liu, Z., 2013. Transit assignment model incorporating bus dwell time. Transp. Res. Rec. 2352 (1), 76–83.
Szeto, W.Y., Jiang, Y., 2014. Transit assignment: approach-based formulation, extragradient method, and paradox. Transp. Res. Part B 62, 51–76.
Szeto, W.Y., Jiang, Y., Wong, K.I., Solayappan, M., 2013. Reliability-based stochastic transit assignment with capacity constraints: formulation and solution method. Transp. Res. Part C 35, 286–304.
Szeto, W.Y., Solayappan, M., Jiang, Y., 2011. Reliability-based transit assignment for congested stochastic transit networks. Computer-Aided Civ. Infrastruct. Eng. 26 (4), 311–326.
Teklu, F., 2008. A stochastic process approach for frequency-based transit assignment with strict capacity constraints. Netw. Spatial Econ. 8 (2-3), 225–240.
Tong, C.O., Wong, S.C., 1999. A stochastic transit assignment model using a dynamic schedule-based network. Transp. Res. Part B 33 (2), 107–121.
Watling, D.P., Rasmussen, T.K., Prato, C.G., Nielsen, O.A., 2015. Stochastic user equilibrium with equilibrated choice sets: part I–Model formulations under alternative distributions and restrictions. Transp. Res. Part B 77, 166–181.
Wu, J., Florian, M., Marcotte, P., 1994. Transit equilibrium assignment: a model and solution algorithms. Transp. Sci. 28 (3), 193–203.
Xue, Y., Jin, J., Lai, J., Ran, B., Yang, D., 2011. Empirical characteristics of transit travel time distribution for commuting routes. Presented. Transportation Research Board 90th Annual Meeting No. 11–2827.
Yang, H., Bell, M.G.H., 1998. A capacity paradox in network design and how to avoid it. Transp. Res. Part A 32 (7), 539–545.
Yang, H., Lam, W.H.K., 2006. Probit-type reliability-based transit network assignment. Transp. Res. Rec. 1977 (1) 154–163.
Zhang, Y., Lam, W.H.K., Sumalee, A., Lo, H., Tong, C.O., 2010. The multi-class schedule-based transit assignment model under network uncertainties. Public Transport 2 (1-2), 69–86.