RELIABILITY-BASED STOCHASTIC TRANSIT ASSIGNMENT WITH CAPACITY CONSTRAINTS: FORMULATION AND SOLUTION METHOD

W.Y. Szeto [1], Yu Jiang [1], K.I. Wong [2], Muthu Solayappan [3]

[1] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, PR China

[2] Department of Transportation Technology and Management. National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu, 30010, Taiwan

[3] Department of Industrial and Systems Engineering, National University of Singapore, 1, Engineering Drive 2, Singapore 117576

ABSTRACT: This paper proposes a Linear Complementarity Problem (LCP) formulation for the reliability-based stochastic transit assignment problem with capacity constraints and non-additive link costs, where in-vehicle travel times and waiting times are uncertain. The capacity constraints are developed via the notions of effective capacity and chance constraints. An equivalent route-based linear program (LP) for the proposed problem is formulated to determine the patronage of each line section, critical links, critical service frequencies, unmet demand and the network capacity, which considers the risk-aversive behaviour of travellers. A solution method is developed, utilizing the K-shortest path algorithm, the column generation technique, and the revised simplex method, to solve the proposed LP with guaranteed finite convergence. Numerical experiments are also set up to illustrate the properties of the problem and the application of the proposed model for reliability analysis.

1. INTRODUCTION

The planning of urban transit services relies on the use of transit assignment models for predicting the way in which transit travellers choose routes from their origins to their destinations (Cepeda *et al*., 2006). As a result, transit assignment models have received considerable attentions over the last two decades. Earliest models (e.g., Dial, 1967; Fearnside and Draper, 1971; Le Clercq, 1972; Nguyen and Pallottino, 1988; De Cea and Fernández, 1989; Spiess and Florian, 1989) assumed that all transit lines have unlimited capacity to accommodate any amount of transit demand, and in-vehicle congestion effects over transit networks were not considered. This limitation does not permit an accurate representation of transit networks with high congestion levels due to insufficient capacity of services (De Cea and Fernández, 1993).

To deal with capacity-related congestion, two approaches have been developed in the literature: the congestion cost function approach and the capacity constraint approach. The congestion cost function approach (e.g., De Cea and Fernández, 1993; Wu *et al.* 1994; Bouzaïene-Ayari *et al.* 1995; Cominetti and Correa, 2001) adopts an unbounded increasing convex function to model the effect of in-vehicle congestion due to insufficient capacity on waiting time. The concept of effective frequency is often used together with this approach. Earlier models such as De Cea and Fernández (1993), Wu *et al.* (1994), and Bouzaïene-Ayari *et al.* (1995) allowed the development of convergent solution algorithms under some monotonicity or uniqueness conditions, but these models allow the flow on a link to be

greater than its capacity, which is unrealistic. The model of Cominetti and Correa (2001) rectified this problem by introducing queue theoretic models, but the algorithm was not necessary convergent.

The capacity constraint approach (e.g., Lam *et al.*, 1999, 2002; Nguyen *et al.*, 2001; Yin *et al.*, 2003; Poon *et al.*, 2003, 2004; Cepeda *et al.*, 2006; Yang and Lam, 2006; Tian *et al.*, 2007a; Teklu, 2008) traditionally incorporates capacity constraints in transit assignment models to disallow the flow on a link to be greater than the corresponding capacity. In the frequency-based method, the excess flow is usually assigned either to a pedestrian arc with unlimited capacity connecting to the destination directly (e.g., Cepeda *et al.*, 2006) or to a failure node (e.g., Kurauchi, *et al.*, 2003; Schmöcker *et al.*, 2008, 2011). In the schedule-based method, the passengers who fail to board a vehicle are kept waiting until they can board the next arriving vehicle with sufficient capacity (e.g., Poon *et al.*, 2004; Hamdouch *et al.*, 2008). Recently, seat capacity has even been considered in the models of Tian *et al.* (2007b), Sumalee *et al.* (2009), Leurent and Liu (2009), Schmöcker *et al.* (2011) and Leurent (2011). This capacity constraint approach is more realistic. However, the resultant models are usually solved by the method of successive averages, which can guarantee convergence under specific conditions. However, these conditions may not be fulfilled by the models.

Other than the congestion effect, most existing transit assignment models do not consider network stochasticity. In fact, due to supply side uncertainty, in-vehicle travel times and waiting times, especially for buses and mini-buses, are highly uncertain. Moreover, most existing transit assignment models do not consider the influence of the variances of travel times on route choice. Indeed, empirical studies like Abdel-Aty *et al.* (1997) and Jackson and Jucker (1982) point out that travel time variability, which is one of the measures of travel time reliability, plays a major role in influencing the trip makers' route choice behaviour. The trip makers select their routes by considering the trade-off between travel time (or cost) and its uncertainty (Yin *et al.*, 2004). It is essential to capture this realistic travel behaviour into the transit modelling framework. To our best knowledge, only Yang and Lam (2006), Li *et al.* (2008, 2009), Szeto and Solayappan (2009a, b), Sumalee *et al.* (2011), and Szeto *et al.* (2011) considered this behaviour in their models. These studies adopted the congestion cost function approach to model the effect of congestion. However, no convergent solution techniques have been developed for their models. Moreover, the capacity constraint approach has not been considered simultaneously with the uncertainties of in-vehicle travel time and waiting time.

This paper proposes a stochastic transit assignment problem with capacity constraints that takes into account the variabilities of in-vehicle travel times and waiting times. In our model, both in-vehicle travel times and waiting times are modelled as random variables. Their means and variances are incorporated in the modelling framework through the concepts of effective travel cost and reliability-based user equilibrium such that the network uncertainty and risk-taking behaviour (including risk-aversive behaviour) of travellers can be captured. The capacity constraints are developed by the chance constraints, which are formulated based on the notion of effective capacity introduced in this paper.

The proposed problem is formulated as a Linear Complementarity Problem (LCP), which is later reformulated as a route-based linear programming problem. This reformulation allows us to determine critical links, critical service frequencies, met and unmet demand, and the transit network capacity. The network capacity considers the risk-averse behaviour of travellers, and it is not required to have any reference Origin-Destination (OD) matrix for determining the capacity. This capacity contrasts to the classical maximum network capacity (e.g., Ahuja *et al.*, 1993) that only considers one OD pair and ignores the route choice behaviour of travellers. Our proposed network capacity also contrasts to the definition of network capacity that relies on a reference OD matrix (e.g., Asakura, 1992; Wong and Yang, 1997; Morlok and Riddle, 1999; Liu and Wei, 2010). Moreover, our proposed capacity differs

from those considering route choice behaviour of travellers (Asakura, 1992; Ahuja *et al*., 1993; Akamatsu and Miyawaki, 1995; Wong and Yang, 1997; Chen *et al.*, 1999; Yang *et al*., 2000; Ziyou and Yifan, 2002; Ge *et al.*, 2003; Kasikitwiwat and Chen, 2005; Lee *et al*., 2006) in the sense that the risk averse behaviour of travellers is taken into account.

This paper proposes a new convergent solution approach, utilizing the column generation technique, the K-shortest path algorithm, and the revised simplex method, to solve the proposed linear programming problem. Numerical examples are also set up to illustrate the properties of the proposed problem and the application of the proposed model for reliability analysis.

The rest of the paper is organized as follows. Section 2 describes the general representation of transit networks, the particular network used for the illustration, and the assumptions that are relevant to our work. Section 3 elucidates the various cost components of the travel cost. Section 4 describes the idea of effective travel costs, and Section 5 postulates the problem formulation. The proposed solution method is detailed in Section 6, whereas Section 7 discusses the computational results. Finally, Section 8 provides concluding remarks and identifies directions for future research.

2. NETWORK REPRESENTATION AND ASSUMPTIONS

Following the method in Lam *et al.* (1999), a general transit network is represented by a set of transit lines and a set of stations (nodes) where passengers can board, alight or transfer. For the simplicity of presentation, a walk link is also represented as a transit line with high frequency. To handle the common line problem and reduce the number of paths handled by the model proposed in Section 3, the line-node transit network is transformed into a link-node network.

To facilitate the discussion, a transit network adopted for the purpose of analysis is shown in Figure 1, which approximates the existing bus network in Singapore. JE, BL, HF, TP, and EU denote five nodes, namely Jurong East, Boon Lay, Harbour Front, Toa Payoh, and Eunos, respectively. Figure 2 shows the same network coded by links (or route sections). In this example, there are four origin-destination (OD) pairs (with HF being a transfer hub) connected by ten routes, and the ten routes are serviced by nine different transit lines. The details are given in Table 1. The fleet sizes of the transit lines as well as the means, variances and covariances of the in-vehicle travel times of the line segments of each line are given in Table 2.

As in the literature (e.g., Spiess and Florian, 1989, De Cea and Fernández, 1993, Szeto *et al.*, 2011), the following classical assumptions are made throughout this paper. A1) Passengers are assumed to arrive at transit stops at a random time. A2) A passenger waiting at a transfer node considers an attractive set of lines before boarding. The set of attractive lines can be determined via the method in Chriqui and Robillard (1975). A3) The waiting time for a transit line on a link is independent of other lines on the same link. A4) Stochastic vehicle headways with the same distribution function are assumed for vehicles servicing different lines. However, different vehicle headways for different lines could be achieved by varying the parameters of the distribution function. A5) A passenger boards the first arriving bus if possible. A6) The passenger knows the mean and variance of in-vehicle travel time of each line (see Szeto *et al.*, 2011). A7) The passenger selects the transit route that minimizes his/her effective travel cost. A8) 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 the day-to-day variation is small and neglectable. A9) Time and cost are used interchangeably throughout this paper by assuming that the value of time is equal to $1 per minute. This can be generalized easily by incorporating the value of time into the model. A10) For simplicity,

the capacity of each transit vehicle is assumed to be the same. However, there is no conceptual difficulty in modelling a scenario in which vehicles of different capacities traverse on different routes.

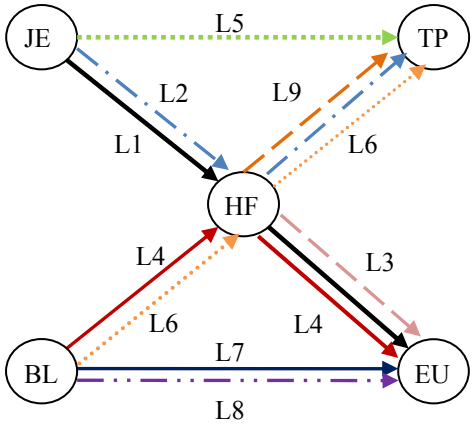

Figure 1. Transit network representation using transit lines

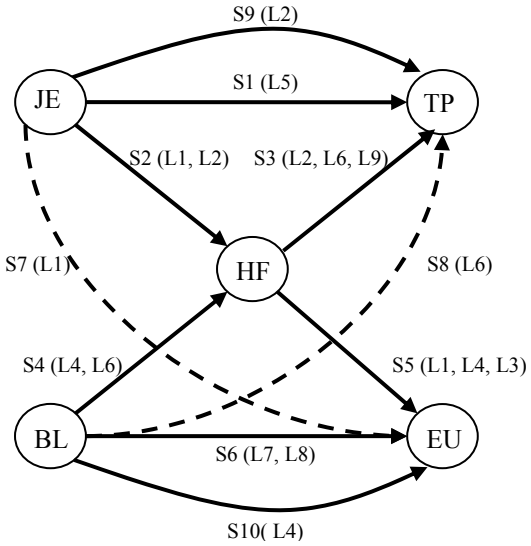

Figure 2. Transit network representation using links

Table 1. Transit route and links

O-D Pairs		Transit Routes		Links (Transit Lines)
1	JE-EU	R1	JE-HF-EU	S2(L1, L2), S5(L1, L3, L4)
		R2	JE-EU	S7(L1)
2	JE-TP	R3	JE-TP	S1(L5)
		R4	JE-TP	S9(L2)
		R5	JE-HF-TP	S2(L1, L2), S3(L2, L6, L9)
3	BL-TP	R6	BL-HF-TP	S4(L4, L6), S3(L2, L6, L9)
		R7	BL-TP	S8(L6)
4	BL-EU	R8	BL-EU	S6(L7, L8)
		R9	BL-EU	S10(L4)
		R10	BL-EU	S4(L4, L6), S5(L1, L3, L4)

Table 2. Fleet sizes, and the mean, variance and covariance of in-vehicle travel times

Transit Line	L1	L2	L3	L4	L5	L6	L7	L8	L9
Fleet size (veh)	18	22	10	20	16	25	18	12	14
Mean (min)	44+45	35+34	37	41+32	65	55+35	70	75	37
Variance (min^2)	6+8	4+3	4	7+6	12	8+11	10	8	6
Covariance (min^2)	2	3	-	2	-	1	-	-	-

Moreover, the following basic assumptions are made. A11) The in-vehicle travel time on a link is subject to randomness due to supply uncertainty. A12) Different buses travelling on the same link can have different scales of travel time variability on that link and their travel times are independent of each other because they can travel on different roads between the two transfer nodes. However, the travel times of the same lines on different links are

dependent because a delay on a link will lead to a delay on a subsequent link. A13) The headway between transit vehicles is exponentially distributed with mean α/f^l, where f^l is defined as the expected frequency of line l, and α is a constant for unit conversion. This assumption is very close to the classical one except that the expected frequency is used to define the headway distribution. Expected frequency is used to model the relationship between stochastic frequency and stochastic in-vehicle travel time. A14) The dwell times are assumed to be constants, which is reasonable when the variabilities of boarding and alighting flows are small. This assumption allows for the development of the formulation that can be solved by convergent algorithms without restrictive assumptions on path cost functions. A15) Unlike Szeto *et al.* (2011), we assume passengers consider the transfer penalty, which is more realistic.

3. INDIVIDUAL COST COMPONENTS

Link cost consists of in-vehicle travel time, the waiting time for the first arriving vehicle, and the passenger overload delay due to insufficient capacity. The following subsections describe these individual cost components.

3.1 In-vehicle Travel Time

Let the random variable T_s^l be the in-vehicle travel time for line l on link s. Then, the weighted average of T_s^l of all the lines on link s is the average in-vehicle travel time on link s, T_s:

$$T_s = \sum_{l \in A_s} \pi_s^l T_s^l, \quad \forall s \in \mathcal{S}, \tag{1}$$

where A_s is the set of attractive lines associated with link s. π_s^l is the relative frequency defined as follows:

$$\pi_s^l = f^l / f_s, \forall l \in A_s, \forall s \in \mathcal{S}. \tag{2}$$

$f^l = E[F^l]$ is the expected frequency of line l; F^l is the service frequency of transit line l; $f_s = \sum_{l \in A_s} f^l$, and $\mathcal{S}$ is the set of links.

Since the expected service frequency is known, the expected in-vehicle travel time can be obtained by taking expectation on both sides of Eq. (1):

$$E[T_s] = \sum_{l \in A_s} \pi_s^l E[T_s^l], \quad \forall s \in \mathcal{S}. \tag{3}$$

Due to assumption A12, $Var[T_s]$ and $Cov[T_s, T_{s'}]$ can be respectively expressed as

$$Var[T_s] = \sum_{l \in A_s} \left(\pi_s^l\right)^2 Var[T_s^l], \quad \forall s \in \mathcal{S}, \text{ and} \tag{4}$$

$$Cov\left[T_s, T_{s'}\right] = \sum_{l \in A_s \cap A_{s'}} \pi_s^l \pi_{s'}^l Cov\left[T_s^l, T_{s'}^l\right], \quad \forall s, s' \in \mathcal{S}. \tag{5}$$

f^l in (2) can be derived as follows. Let the random variables F^l, N^l, and Γ^l be the line frequency, fleet size and round trip time of transit line $l \in \mathcal{L}$ respectively, where $\mathcal{L}$ is the set of transit lines. Then, by definition, $F^l = N^l / \Gamma^l$. By Li *et al.* (2008), we have

$$f^l = E\left[F^l\right] = \frac{N^l}{E\left[\Gamma^l\right]}\left(1+\frac{Var\left[\Gamma^l\right]}{E^2\left[\Gamma^l\right]}\right), \quad \forall l \in \mathcal{L}. \tag{6}$$

By definition, $\Gamma^l = \xi^l t_0^l + n^l t_1^l + \sum_{s\in\mathcal{S}^l} T_s^l$, where ξ^l is an indicator variable, which equals 1 when line l is a circular line and 2 otherwise. t_0^l is the layover time of transit line l; t_1^l and n^l are respectively the dwell time and the number of line segments in the line-node network visited in the round trip of transit line l. $\mathcal{S}^l$ is the set of links in the link-node network corresponding to the set of line segments in the line-node network that are visited by transit line l in the round trip. Hence, $E\left[\Gamma^l\right]$ and $Var\left[\Gamma^l\right]$ in (6) can be expressed as

$$E\left[\Gamma^l\right] = \xi^l t_0^l + n^l t_1^l + \sum_{s\in\mathcal{S}^l} E\left[T_s^l\right], \quad \forall l \in \mathcal{L}, \text{ and} \tag{7}$$

$$Var\left[\Gamma^l\right] = \sum_{s\in\mathcal{S}^l} Var[T_s^l] + \sum_{s\in\mathcal{S}^l}\sum_{s'\neq s, s'\in\mathcal{S}^l} Cov\left[T_s^l, T_{s'}^l\right], \quad \forall l \in \mathcal{L}. \tag{8}$$

3.2 Waiting Time

Waiting time on a link is defined as the time a passenger waits at a transit stop (the origin node of the link) for the arrival of the *first* vehicle belonging to the attractive set considered by the passenger. Under assumptions A1-A4 and A13, the mean and variance of the random waiting time X_s on link s can be derived as (see Szeto *et al.*, 2011):

$$E[X_s] = \frac{\alpha}{f_s}, \quad \forall s \in \mathcal{S}, \text{ and} \tag{9}$$

$$Var(X_s) = E^2[X_s], \quad \forall s \in \mathcal{S}. \tag{10}$$

where α is a constant for unit conversion.

3.3 Passenger overload delay

Due to the limited capacity of each transit vehicle, passengers may not be able to board the first arriving vehicle at a station (the origin node of a link) and may experience *extra* delay due to passenger overload. Hence, it is important to consider the effects of passenger overload delay. When congestion exists in a transit network, it is difficult (if possible) to derive an analytical formula to model the overload delay due to the existence of various stochastic effects of both passenger behaviour and overcrowded vehicle arrivals (Lam *et al.*, 1999). Hence, the passenger overload delay in a congested transit network is determined endogenously and described by an equilibrium condition as shown below.

The random capacity K_s on link s is given by

$$K_s = \frac{\gamma k}{H_s}, \quad \forall s \in \mathcal{S}, \tag{11}$$

where k is the capacity of the vehicle (in passengers/vehicle); H_s is the random headway of vehicles servicing link s; and γ is a unit conversion factor. γ = 60 min/h if the unit for headway is minutes and the unit for the capacity of a line is passengers per hour.

Equation (11) implies that each headway value h_s is related to one and only one capacity value k_s and vice versa. Therefore, we have

$$h_s = \frac{\gamma k}{k_s}, \quad \forall s \in \mathcal{S}, \text{ and} \tag{12}$$

$$\frac{dh_s}{dk_s} = -\frac{\gamma k}{k_s^2}. \tag{13}$$

Based on the superposition of Poisson processes for all the bus lines on link s, the random headway H_s in (11) follows an exponential distribution:

$$f_{H_s}(h_s) = \frac{f_s}{\alpha} e^{-\frac{f_s h_s}{\alpha}}, \quad \forall s \in \mathcal{S}. \tag{14}$$

Since H_s is a continuous random variable and (11) defines a one-to-one correspondence between the values of H_s and K_s, according to Proposition 7.3 in Walpole *et al.* (2007), the density function of the random capacity K_s of link s can be obtained by

$$f_{K_s}(k_s) = \left|\frac{dh_s}{dk_s}\right| f_{H_s}\left(\frac{\gamma k}{k_s}\right), \quad \forall s \in \mathcal{S}, \tag{15}$$

By substituting (13)-(14) into (15), we get the density function of K_s:

$$f_{K_s}(k_s) = \frac{\gamma k}{k_s^2} \frac{f_s}{\alpha} e^{-\frac{f_s}{\alpha}\frac{\gamma k}{k_s}}, \quad \forall s \in \mathcal{S}. \tag{16}$$

The distribution function of K_s is therefore equal to

$$F(K_s \le \varphi_s) = \int_0^{\varphi_s} \frac{\gamma k}{k_s^2} \frac{f_s}{\alpha} e^{-\frac{f_s}{\alpha}\frac{\gamma k}{k_s}} dk_s = e^{-\frac{f_s}{\alpha}\frac{\gamma k}{\varphi_s}}, \quad \forall s \in \mathcal{S}, \tag{17}$$

where φ_s is the effective link flow on s, which can be defined as

$$\varphi_s = v_s + \tilde{v}_s = \sum_{l \in A_s} v_s^l + \sum_{m \in \mathcal{S}} \sum_{l \in A_s \cap A_m} \delta_{sm} v_m^l, \forall s \in \mathcal{S}, \tag{18}$$

where v_s is the flow on link *s*; $\tilde{v}_s$ is the total flow on the competing links of *s*; v_s^l is the flow of line *l* on link *s*; δ_{sm} is an indicator variable, which equals 1 if link *m* competes with link *s*, and 0 otherwise (In particular, $\delta_{ss} = 0$). Link *m* is a competing link of *s* if link *m* contains at least one attractive line of link *s* and one of the two conditions is satisfied: 1) The origin node of *m* is before that of *s* and the destination node of *m* is after the origin node of *s*, or 2) the origin node of *m* is the same as *s* but the destination nodes of the two links are not the same. This effective link flow captures the interaction between links and the passengers on different links who compete for the residual capacity of the same set of attractive lines.

Like De Cea and Fernández (1993), v_s^l in (18) can be calculated through

$$v_s^l = \pi_s^l v_s, l \in A_s, \forall s \in \mathcal{S}. \tag{19}$$

By using the chance constraint approach, a relation between the *effective capacity* and the effective link flow φ_s can be derived. Let α_s be the maximum violation probability of link s with $0 < \alpha_s \le 1$, which is the largest probability that the flow on link s can be greater than or equal to the link capacity. α_s is a parameter defined by the modeler. A smaller value of α_s implies a more stringent requirement on the link flow. Then,

$$F(K_s \le \varphi_s) = e^{-\frac{f_s}{\alpha}\frac{\gamma k}{\varphi_s}} \le \alpha_s, \quad \forall s \in \mathcal{S}. \tag{20}$$

By rearranging (20), we get the capacity constraint:

$$-\frac{\gamma k f_s}{\alpha \ln \alpha_s} \geq \varphi_s, \quad \forall s \in \mathcal{S}. \tag{21}$$

The term on the left hand side of (21) is referred to as the effective capacity of link s, and is nonnegative since $\ln \alpha_s$ is always non-positive for $0 < \alpha_s \leq 1$.

Based on (21), the equilibrium conditions for passenger overload delay on link *s* can be defined as follows. The passenger overload delay, d_s, is positive if the effective link flow φ_s on link *s* is greater than the effective link capacity given on the left hand side of Eq. (21), and equals zero if the effective link flow is less than or equal to the effective link capacity. The rationale is that when there is insufficient capacity in the first arriving vehicle, some passengers cannot board this vehicle, thereby, causing delays. Mathematically, the equilibrium condition can be expressed as follows:

$$d_s \geq 0, \quad \forall s \in \mathcal{S}, \tag{22}$$

$$d_s \left[-\varphi_s - \frac{\gamma k f_s}{\alpha \ln \alpha_s} \right] = 0, \quad \forall s \in \mathcal{S}, \text{ and} \tag{23}$$

$$-\varphi_s - \frac{\gamma k f_s}{\alpha \ln \alpha_s} \geq 0, \quad \forall s \in \mathcal{S}. \tag{24}$$

Due to sharing the capacity of the same set of lines between passengers on different links, the flow on link *s* experiences overload delay due to its own link, d_s, as well as overload delay due to competing link *m*. The overload delay of link *s* from competing link *m* is the overload delay of the competing link d_m times the proportion of the capacity of link *s* shared by the two links, $\sum_{l \in A_s \cap A_m} \pi_s^l$. The total overload delay of link *s* from all competing links *m* is therefore $\sum_{m \in S} \delta_{sm} d_m \sum_{l \in A_s \cap A_m} \pi_s^l$, where $\delta_{sm} = 1$ if link *m* competes with link *s*, and $\delta_{ms} = 0$ otherwise. Hence, the total overload delay on link *s* is $\tilde{d}_s = d_s + \sum_{m \in S} \delta_{sm} d_m \sum_{l \in A_s \cap A_m} \pi_s^l$. This total overload delay is the extra waiting time for the passengers boarding at the origin node of link s.

4. EFFECTIVE TRAVEL COST

The variabilities associated with the in-vehicle travel time and waiting time, along with the delay due to congestion, causes variability in trip time. Consequently, a passenger cannot determine the exact trip time to complete his/her journey. The passenger counters the variability in trip time by an early departure to allow for additional time for the trip and avoid being late. The additional time is referred to as the safety margin, and depends on both the purpose of the trip and the individual's risk taking behaviour. This safety margin plus the expected trip time is the effective travel cost, which has been applied in Lo *et al.* (2006), Shao *et al.* (2006), Lam *et al.* (2008), Siu and Lo (2008), and Szeto *et al.* (2012). Mathematically, the effective travel cost associated with route *r* between OD pair *w*, η_r^w, can be expressed as:

$$\eta_r^w = E[C_r^w] + \rho \sqrt{Var[C_r^w]}, \ \forall r \in \mathcal{R}^w, w \in \mathcal{W}, \tag{25}$$

where C_r^w is the trip time (including total in-vehicle travel time and total waiting time for the first arriving vehicle for the trip) on route r connecting OD pair w and is a random variable.

$\mathcal{R}^w$ denotes the set of routes connecting OD pair w, the set of which is denoted by $\mathcal{W}$. ρ is the parameter representing the degree of risk aversion of passengers. A higher value of ρ means a more risk-aversive passenger and leads to a larger safety margin, $\rho\sqrt{Var[C_r^w]}$.

According to Lo *et al.* (2006), the parameter ρ relates to the probability λ that the actual trip time is less than the effective travel cost:

$$P\{C_r^w \le \eta_r^w = E[C_r^w] + \rho\sqrt{Var[C_r^w]}\} = \lambda, \ \forall r \in \mathcal{R}^w, w \in \mathcal{W}. \tag{26}$$

This probability can be regarded as the within cost budget reliability, in which the cost budget is defined by the effective travel cost. Then, a higher value of ρ implies that the passenger is willing to have a higher probability of arriving on time or equivalently the actual travel cost not being greater than the actual effective travel cost.

The route cost in (25) is the sum of the total transfer penalty cost, the total dwell time, and the costs of links on the route. The total penalty cost represents the discomfort cost due to the transfer inconvenience. Let $\tilde{P}$ be the penalty cost for each transfer, and b_{sr} be the link-path incidence indicator variable, which equals 1 if link s is a part of transit route (path) r, and equals 0 otherwise. Then, the total transfer penalty on route r is $\tilde{P}_r = \left(\sum_{s\in S} b_{sr} - 1\right)\tilde{P}$.

The dwell time at the tail node of the link is the weighted average of the dwell time $\sum_{l\in A_s} \pi_s^l t_1^l$. Hence, the total dwell time on route r is $\sum_{s\in\mathcal{S}} b_{sr} \sum_{l\in A_s} \pi_s^l t_1^l$.

The link cost, C_s, is the sum of in-vehicle travel time, waiting time and total delay due to passenger overload. That is, $C_s = T_s + X_s + \tilde{d}_s$.

The cost on route r can be expressed as

$$C_r^w = \sum_{s\in S} b_{sr}\left[T_s + X_s + \tilde{d}_s\right] + \tilde{P}_r + \sum_{s\in\mathcal{S}} b_{sr} \sum_{l\in A_s} \pi_s^l t_1^l, \ \forall r \in \mathcal{R}^w, w \in \mathcal{W}. \tag{27}$$

By taking the expectation and variance on both sides of Eq. (27) and substituting the two resulting expressions into Eq. (25), we get

$$\begin{aligned} \eta_r^w = {} & \sum_{s\in\mathcal{S}} b_{sr}\left(E[T_s] + E[X_s] + \tilde{d}_s\right) + \tilde{P}_r + \sum_{s\in\mathcal{S}} b_{sr} \sum_{l\in A_s} \pi_s^l t_1^l \\ & + \rho\sqrt{\sum_{s\in\mathcal{S}} b_{sr}\left(Var[T_s] + Var[X_s]\right) + \sum_{s\in\mathcal{S}} \sum_{s'\ne s, s'\in\mathcal{S}} b_{sr} b_{s'r} Cov\left[T_s, T_{s'}\right]}, \forall r \in \mathcal{R}^w, w \in \mathcal{W}. \end{aligned} \tag{28}$$

This route cost can be decomposed into the effective uncongested travel cost $\tilde{\eta}_r^w$ and the route's overload delay d_r^w. Hence, $\eta_r^w = d_r^w + \tilde{\eta}_r^w$ where $d_r^w = \sum_{s\in S} b_{sr}\tilde{d}_s$, and

$$\begin{aligned} \tilde{\eta}_r^w = {} & \sum_{s\in S} b_{sr}\left(E[T_s] + E[X_s]\right) + \left(\sum_{s\in S} b_{sr} - 1\right)\tilde{P} + \sum_{s\in\mathcal{S}} b_{sr} \sum_{l\in A_s} \pi_s^l t_1^l \\ & + \rho\sqrt{\sum_{s\in S} b_{sr}\left(Var[T_s] + Var[X_s]\right) + \sum_{s\in\mathcal{S}} \sum_{s'\ne s, s'\in\mathcal{S}} b_{sr} b_{s'r} Cov\left[T_s, T_{s'}\right]}, \forall r \in \mathcal{R}^w, w \in \mathcal{W}. \end{aligned} \tag{29}$$

5. RELIABILITY-BASED STOCHASTIC TRANSIT ASSIGNMENT FORMULATION

5.1 Linear Complementarity Problem Formulation

Under assumption A7, we can define the reliability-based user equilibrium (RUE) as follows: *The transit network is said to be at RUE for each OD pair, if the effective travel cost of routes*

with non-zero flows are equal to each other and not greater than that of any route with no flow. The RUE conditions can mathematically be stated as follows:

$$y_r^w \geq 0, \quad \forall r \in \mathcal{R}^w, \ w \in \mathcal{W}, \tag{30}$$

$$\eta_r^w - u^w \geq 0, \quad \forall r \in \mathcal{R}^w, \ w \in \mathcal{W}, \text{ and} \tag{31}$$

$$y_r^w (\eta_r^w - u^w) = 0, \quad \forall r \in \mathcal{R}^w, \ w \in \mathcal{W}, \tag{32}$$

where u^w is the reliability-based equilibrium travel cost over all the routes that connect OD pair $w \in \mathcal{W}$ and y_r^w is the passenger flow on route $r \in \mathcal{R}^w$.

Apart from the above RUE conditions (30)-(32), the overload delay conditions (18)-(19), (22)-(24), the effective travel cost condition (28), and the conditions of the means and covariances of in-vehicle travel and waiting times (2)-(10), the proposed problem includes the flow conservation constraints and the relationship between link flows and route flows:

$$\sum_{r \in \mathcal{R}^w} y_r^w - q^w = 0, \quad \forall w \in \mathcal{W}, \text{ and} \tag{33}$$

$$v_s = \sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}^w} b_{sr} y_r^w, \quad \forall s \in \mathcal{S}, \tag{34}$$

where q^w is the demand for OD pair $w \in \mathcal{W}$.

In fact, (33) can be reformulated into the complementarity condition:

$$u^w \geq 0, \quad \forall w \in \mathcal{W}, \tag{35}$$

$$u^w \left(\sum_{r \in \mathcal{R}^w} y_r^w - q^w \right) = 0, \quad \forall w \in \mathcal{W}, \text{ and} \tag{36}$$

$$\sum_{r \in \mathcal{R}^w} y_r^w - q^w \geq 0, \quad \forall w \in \mathcal{W}, \tag{37}$$

because at optimality, $u^w > 0$, and hence (33) must hold. Then, the proposed problem can be expressed as a Linear Complementarity Problem (LCP).

The LCP is to find $\mathbf{Z} = \left[\mathbf{Y}^T, \mathbf{D}^T, \mathbf{U}^T \right]^T$ such that,

$$\mathbf{Z} \geq \mathbf{0}, \text{ and} \tag{38}$$

$$\mathbf{MZ + V} \geq \mathbf{0}, \ \mathbf{Z}^T \left(\mathbf{MZ + V} \right) = 0, \tag{39}$$

where $\mathbf{Y} = \left[y_r^w \right]$ with dimension $|\mathcal{P}| = \sum_{w \in \mathcal{W}} |\mathcal{R}^w|$; $\mathbf{D} = \left[d_s \right]$ with dimension $|\mathcal{S}|$; $\mathbf{U} = \left[u^w \right]$ with dimension $|\mathcal{W}|$; $\mathbf{M} = \begin{bmatrix} \mathbf{0} & \mathbf{B} & -\mathbf{A} \\ -\mathbf{H} & \mathbf{0} & \mathbf{0} \\ \mathbf{A^T} & \mathbf{0} & \mathbf{0} \end{bmatrix}$ with dimension $\left(|\mathcal{P}| + |\mathcal{S}| + |\mathcal{W}| \right) \times \left(|\mathcal{P}| + |\mathcal{S}| + |\mathcal{W}| \right)$; $\mathbf{V} = \left[\mathbf{C}^T, -\mathbf{F}^T, -\mathbf{Q}^T \right]^T$ with dimension $\left(|\mathcal{P}| + |\mathcal{S}| + |\mathcal{W}| \right)$; $\mathbf{B} = \left[b_{sr} \right]$ with dimension $|\mathcal{P}| \times |\mathcal{S}|$; $\mathbf{A} = \left[a_r^w \right]$ with dimension $|\mathcal{P}| \times |\mathcal{W}|$, with $a_r^w = 1$ if route r connects OD pair w and 0 otherwise; $\mathbf{H}$ is a matrix with $|\mathcal{S}| \times |\mathcal{P}|$ such that $\mathbf{HY} = \left[\varphi_s \right]$ with dimension $|\mathcal{S}| \times 1$ or equivalently $\mathbf{HY} = \left(\mathbf{G} \left(\left(\mathbf{E} diag(\mathbf{f}) \right) \circ \left\{ \left(\mathbf{B^T Y} \left(\frac{1}{\mathbf{Ef}} \right) \right) \cdot \mathbf{1}_L^T \right\} \right) \right) \circ \mathbf{E} \cdot \mathbf{1}_L + \mathbf{B^T Y}$; $\mathbf{f} = \left[f^l \right]$ with dimension $|\mathcal{L}| \times 1$; $\mathbf{G} = \left[\delta_{sm} \right]$ with dimension $|\mathcal{S}| \times |\mathcal{S}|$; $\mathbf{E} = \left[e_s^l \right]$ with dimension $|\mathcal{S}| \times |\mathcal{L}|$, e_s^l = 1 if line l is attractive on link s, and 0 otherwise; $\mathbf{1}_L = [1]$ with dimension $|\mathcal{L}| \times 1$;

$\mathbf{F} = \left[\frac{\gamma k f_s}{\alpha \ln \alpha_s} \right]$ with dimension $|\mathcal{S}|$; $\mathbf{Q} = \left[q^w \right]$ with dimension $|\mathcal{W}|$; $\mathbf{C} = \left[\tilde{\eta}_r^w \right]$ with dimension $|\mathcal{P}|$ and $\tilde{\eta}_r^w$ follows (29).

5.2 Linear Programming Reformulation

The problem formulation detailed above can be expressed as a Linear Program (LP):

$$\underset{\mathbf{Y}}{\text{Min}} \sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}^w} \tilde{\eta}_r^w y_r^w \tag{40}$$

s.t. the non-negativity constraint (30), the flow conservation constraint (33) and

$$-\left[\sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}^w} b_{sr} y_r^w + \sum_{m \in \mathcal{S}} \sum_{l \in A_s \cap A_m} \delta_{sm} \pi_m^l \sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}^w} b_{mr} y_r^w \right] - \frac{\gamma k f_s}{\alpha \ln \alpha_s} \geq 0, \qquad \forall s \in \mathcal{S}. \tag{41}$$

In this formulation, the objective is to minimize the sum of the effective *uncongested* travel costs of all passengers. Let u^w and d_s be the multipliers attached to constraint (33) and capacity constraint (41), respectively. Then, the first-order conditions can be expressed as (22)-(23), (30)-(33), and (41) where φ_s is defined by (18)-(19) and (34), and η_r^w is defined by (28). In other words, this LP is equivalent to the LCP (38)-(39). u^w and d_s are the dual solution of the LP. In particular, d_s can be interpreted as the rate of the reduction of the total effective *uncongested* travel cost at equilibrium with respect to the increase in the effective capacity of link *s*. It is the decrease in the total effective uncongested travel costs at optimality when one unit of effective capacity is added to link *s*.

The LP can be written in matrix form as shown below:

Problem *P*

$$\text{Min } Z_p = \mathbf{C}^T \mathbf{Y} \tag{42}$$

Subject to $\mathbf{A}^T \mathbf{Y} = \mathbf{Q}$, (43)

$-\mathbf{H} \cdot \mathbf{Y} - \mathbf{F} \geq \mathbf{0}$, and (44)

$\mathbf{Y} \geq \mathbf{0}$. (45)

Problem *P* may not have a solution because the effective capacity of certain links may be insufficient to handle a large demand. In case we need to know which link has insufficient effective capacity, analyzing problem *P* is not enough. To ensure the existence of solutions for analysis, we create an artificial problem using the following assumption.

Assumption 16: There is one direct virtual (or artificial) path connecting each origin and destination with the capacity at least equal to the corresponding OD demand. Each direct virtual path has a huge effective travel cost M, where M is much larger than the effective travel cost on any actual (or real) paths.

Introducing virtual paths with their individual capacity not less than the corresponding demand can ensure solution existence. However, if the effective (uncongested) travel costs on the virtual paths were low, including these paths in the formulation could affect the optimality due to the fact that the virtual paths are more attractive than the actual paths. This will result in a scenario where there are positive flows on the virtual paths even when the network capacity is sufficient. Therefore, the travel costs on all virtual paths must be large enough to make these paths unattractive even when sufficient network capacity exists. These virtual paths can then be interpreted as pedestrian links that the passengers walk to their destinations. The cost on these paths is so huge that the passengers prefer to travel by bus in case the capacity is sufficient.

With the above assumption, a more general problem is stated below:

Problem *P(M)*

$$\text{Min } Z_M = \mathbf{C}^T\mathbf{Y} + M\mathbf{Y}_a \tag{46}$$

subject to (44)-(45)

$$\mathbf{A}^T\mathbf{Y} + \mathbf{Y}_a = \mathbf{Q}, \tag{47}$$

$$\mathbf{Y}_a \geq \mathbf{0}, \tag{48}$$

where $\mathbf{Y}_a$ is the path flow vector for virtual paths. Note that in this problem, the capacity constraint $\mathbf{Y}_a \leq \mathbf{Q}$ for the virtual paths is redundant as the first equality constraint (47) and two non-negativity constraints (45) and (48) imply the capacity constraints for virtual paths.

Problem *P(M)* has the following properties:

(i) *P(M)* has at least one optimal solution (see propositions 1 and 2 in the Appendix).

(ii) If there are no flows on all virtual paths at optimality, the flows on actual paths form an optimal solution for problem *P*. Otherwise, there is no feasible solution for problem *P* (see proposition 3 in the Appendix), and

(iii) The maximum possible number of used paths in problem *P(M)* is $|\mathcal{S}|+|\mathcal{W}|$ (see proposition 4 in the Appendix).

Property (ii) implies that problem *P(M)* includes problem *P* as a special case. To obtain optimal solutions of *P* or to show that *P* has no feasible solution, we can solve *P(M)* instead. Property (iii) implies that for large networks, most paths carry no flows. This property has an important implication on developing efficient solution methods. We do not need to have a complete path set for obtaining optimal solutions. Instead we need a systematic procedure to exclude the paths with zero flows at optimality. This leads to the idea of incorporating the column generation procedure in the solution method.

Compared with using the Big-M method to solve problem *P* directly, our proposed method of solving *P(M)* using the column generation method (discussed in the next section) introduces the concept of virtual paths to the proposed transit assignment problem. The notion of virtual paths gives some physical meaning to the artificial variables normally used in the big M-method. The proposed method also gives some useful information like the *unmet demand*, the *met demand*, and the *deficit capacity* of each OD pair, *total met demand*, *network capacity*, *critical links*, and *critical service frequencies* for transit network reliability analysis. The unmet demand of an OD pair *w* is the positive optimal flow y_a^{w*} on the virtual path of that OD pair, where y_a^w is the element of $\mathbf{Y}_a$ and the asterisk denotes the solution at optimality. This unmet demand numerically equals the deficit capacity of that OD pair. To handle the unmet demand, extra capacity (which at least equals the unmet demand) must be added to that OD pair by say, increasing the frequency of the bus line or the capacity of the buses serving that OD pair. The sum of optimal flows on all paths $r \in \mathcal{R}^w$ between OD pair *w* (i.e., $\sum_{r\in\mathcal{R}^w} y_r^{w*}$) gives the met demand of that OD pair, and the sum of the met demand of each OD pair gives the total met demand (i.e., $\sum_{w\in\mathcal{W}}\sum_{r\in\mathcal{R}^w} y_r^{w*}$). In the extreme case, when all virtual paths carry flows at optimality (i.e., $y_a^{w*} > 0, \forall w$), the total met demand gives the network capacity $Q_{\max}$ and is equal to the difference between the total demand and the total unmet demand. Mathematically, the maximum capacity can be represented as $Q_{\max} = \sum_{w\in\mathcal{W}}\sum_{r\in\mathcal{R}^w} y_r^{w*}$ if $y_a^w > 0, \forall w$ or $Q_{\max} = \sum_{w\in\mathcal{W}}\left(q^w - y_a^{w*}\right)$ if $y_a^w > 0, \forall w$. Finally, the critical links can be obtained by finding the links with zero residual capacities or the links with

positive overload delays (i.e. $d_s > 0$) whereas the frequencies on the critical links are critical service frequencies that should be increased to improve the network capacity.

6. SOLUTION METHOD

The proposed problem *P(M)* creates several challenges for obtaining solutions. First, the LP proposed here is path-based and cannot be reformulated as link-based, because the standard deviation of path travel cost is path-specific and not equal to the sum of the standard deviations of the link travel costs on the path. Therefore, the existing link-based LP solution methods cannot be employed for solving this problem. Second, existing path-based solution methods for LP require that path costs are additive of link costs. However, in our problem, the effective path travel costs are not the sum of effective link travel costs. Hence, existing path-based methods cannot be used for solving the proposed problem. Third, path-based formulations normally require a complete path set. For large networks, it is very time consuming to determine all paths. Fourth, the existence of many path flow variables for large networks may pose the computer storage problems and impact the computational speed.

In this paper, we develop a new solution algorithm to solve the proposed problem *P(M)*, which does not require generating a complete path set and storing all route flow variables. This algorithm generates a path only when needed. This algorithm relies on the K-shortest path algorithm and the revised simplex method. The K-shortest path algorithm forms the backbone of the column generation procedure to generate real paths (i.e., non-virtual paths) for the reduced path-based LP at every iteration i, $RP_i(M)$:

Problem $RP_i(M)$

$$\text{Min } Z_i = \mathbf{C}_i^T \mathbf{Y}_i + M\mathbf{Y}_a \tag{49}$$

Subject to

$$\mathbf{A}_i^T \mathbf{Y}_i + \mathbf{Y}_a = \mathbf{Q}, \tag{50}$$

$$-\mathbf{H}_i \ \mathbf{Y}_i - \mathbf{F} \geq \mathbf{0}, \tag{51}$$

$$\mathbf{Y}_i \geq \mathbf{0}, \text{ and} \tag{52}$$

$$\mathbf{Y}_a \geq \mathbf{0}, \tag{53}$$

where $\mathbf{C}_i, \mathbf{Y}_i, \mathbf{A}_i,$ and $\mathbf{H}_i$ correspond to $\mathbf{C}, \mathbf{Y}, \mathbf{A},$ and $\mathbf{H}$ but only include a subset of paths in the whole network, and the i-th row is for the paths added in iteration i. The revised simplex method is used to solve each reduced problem. The proposed algorithm is terminated only when one of the following conditions is satisfied:

(i) For each OD pair where not all paths are generated, the largest mean travel cost of the generated but not included path is larger than the largest effective travel cost found with the paths included so far.

(ii) All paths in the network are generated.

Otherwise, the algorithm selects an OD pair that does not satisfy the above conditions, and then generates a single path that connects the OD pair. The new path is added to the reduced problem to form $RP_{i+1}(M)$ for re-optimization and the solution obtained is checked for convergence again.

The detailed algorithmic steps are as follows:

1. Set $i = 0$, $k^w = 0$, $P^w = 0$, $\tilde{u}_i^w = M$, $\omega_i^w = 0$, $\forall w \in \mathcal{W}$, where $\tilde{u}_i^w$ and ω_i^w are the largest effective uncongested travel cost on the used path between OD pair *w* and the largest mean travel cost of OD pair *w* at iteration *i*, respectively. P^w is an indicator variable, which is equal to 1 if all paths between *w* are generated and 0 otherwise. Let the set of

paths for OD pair w, $\mathcal{R}_i^w = \{a^w\}$, where a^w is the virtual path of OD pair w. Assign the demand of each OD pair on the corresponding virtual paths.

2. If (i) $\omega_i^w > \tilde{u}_i^w, \forall w \in \mathcal{W}$, or (ii) $P^w = 1, \forall w \in \mathcal{W}$, then stop.
3. Select an OD pair w with $P^w = 0$ and $\omega_i^w \le \tilde{u}_i^w$. Set $k^w = k^w + 1$. Solve the k^{rs}-shortest path problem for OD pair w based on the mean link travel cost $E[\tilde{C}_s] = E[T_s] + E[X_s] + \sum_{s \in \mathcal{S}} b_{sr} \sum_{l \in A_s} \pi_s^l t_1^l$ where $E[T_s]$ and $E[X_s]$ are calculated via (3)-(10). If no new real path can be generated, then set $\mathcal{P}^w = 1$, and go to step 2.
4. For new path r, compute $E[\tilde{C}_r^w] = \sum_{s \in S} b_{sr} \left(E[T_s] + E[X_s] \right) + \left(\sum_{s \in S} b_{sr} - 1 \right) \tilde{P} + \sum_{s \in \mathcal{S}} b_{sr} \sum_{l \in A_s} \pi_s^l t_1^l$. Set $\omega_i^w = \max\left\{ E[\tilde{C}_r^w], \omega_{i-1}^w \right\}$ and then determine the effective uncongested travel cost $\tilde{\eta}_r^w$ by (29). Set $i = i + 1$, and $\mathcal{R}_i^w = \mathcal{R}_{i-1}^w \cup \{r\}$. Form *RP$_i$(M)* using $\mathcal{R}_i^w, \forall w \in \mathcal{W}$. Solve *RP$_i$(M)* by the revised simplex method and update $\tilde{u}_i^w, \forall w \in \mathcal{W}$. Go to step 2.

This algorithm adopts the revised simplex method to solve a sequence of reduced problems based on a subset of paths in the original network. To apply the revised simplex method, a feasible initial solution must be provided. The initial solution in the first iteration can be formed by setting each virtual path flow to be equal to the corresponding demand and the flows on the remaining paths to be zeros. The initial solution for the subsequent iterations can be formed by setting the flow of the newly added path to zero and the flows on remaining paths can be set at their optimal values found in the preceding iteration. This initial solution should be closer to the optimal solution for the complete network than the initial solution used in the first iteration.

Over iterations, the algorithm shows the following characteristics:

1. The optimal objective value calculated by the algorithm is non-increasing with every iteration (see proposition 5 in the Appendix).
2. The optimal objective value obtained by the algorithm at the next iteration must be lower than that of the current iteration if the new path added for the next iteration has a positive residual effective capacity before re-optimization, (where the residual effective capacity of the new path is the minimum flow required to cause at least one link on this path to have its link flow equal to its effective capacity) and the virtual path connecting the same OD pair as the new path carries flow before re-optimization (see proposition 6 in the Appendix).
3. The optimal objective value of the next iteration must be lower than that of the current iteration if all the following conditions are simultaneously satisfied before re-optimization: 1) the virtual path connecting the same OD pair as the new path carries flow; and 2) all binding links (links with zero residual effective capacities) on this new path are part of another used route that is longer than the new path (see proposition 7 in the Appendix).
4. The largest mean travel cost for each OD pair is non-decreasing with every iteration (see proposition 8 in the Appendix).

As mentioned before, the algorithm can be terminated under two situations:

1. When the algorithm is terminated due to condition (i), the flows on the used paths in the current reduced problem are the optimal flows on the corresponding used paths in problem *P*, the flows on other paths in problem *P* are zeros, and the current objective value is the optimal objective value (see Proposition 9 in the Appendix).
2. When the algorithm is terminated due to condition (ii), whether an optimal solution is

obtained can be judged by the flows on virtual paths according to Proposition 3. If all these virtual paths carry no flows, an optimal solution is obtained. Otherwise, no solution exists to problem *P*.

The revised simplex method and the K-shortest path algorithm guarantee finite convergence. Moreover, the column generation technique here only allows adding paths and in the worst case, the column generation technique builds all real paths, where the number of paths is finite. Hence, the column generation technique guarantees finite convergence, thereby allowing the proposed algorithm to guarantee finite convergence. Note that even if the maximum possible number of used paths at optimality is large, our algorithm guarantees finite convergence because the convergence proof does not rely on this number.

7. NUMERICAL RESULTS

Five studies will be carried out using the example network discussed in Section 2. The basic link data related to the network is given in Tables 1-2. All transit lines are assumed to be served by the single deck bus, Mercedes Benz O 405, which currently operates in Singapore to serve the entire network. This bus model has a total capacity of 85 passengers. The transfer penalty cost $\tilde{P}$ and the effective travel cost on virtual path are, respectively, set to be 30 and 1000. The maximum violation probability, α_s, is set to 0.05 unless otherwise specified, and the headway is assumed to be exponentially distributed with mean $1/f^l$. The dwell time at each stop is set to 1 minute and the layover time is set to 15 minutes.

7.1 Application to transit network reliability analysis

This study aims to illustrate the applications of the proposed model for transit network reliability analysis. The demand is set to be 500 passengers/hr for each OD pair. A LP was developed for this scenario and solved by the solution method described in Section 6. The optimal results with $\rho = 2.75$ is presented in Tables 3 and 4.

Table 3. Optimal solutions and effective travel cost

w	q^w	u^w	r	y_r^w	E_r^w	$\eta_r^w - \tilde{\eta}_r^w$
1	500.0	1000.0	1	0.0	1021.5	887.2
			2	144.7	1000.0	862.8
2	500.0	1000.0	3	168.3	1000.0	894.5
			4	217.9	1000.0	897.5
			5	0.0	1018.2	887.2
3	500.0	1000.0	6	0.0	1019.2	880.6
			7	199.1	1000.0	872.8
4	500.0	1000.0	8	290.6	1000.0	904.0
			9	189.4	1000.0	888.8
			10	0.0	1022.9	880.6

Table 4 Residual capacity and overload delay

s	d_s	Residual capacity
1	894.5	0.0
2	862.8	0.0
3	0.0	225.1
4	872.8	0.0
5	0.0	160.7
6	904.0	0.0
7	0.0	0.0
8	0.0	0.0
9	34.7	0.0
10	16.0	0.0

According to Table 3, the equilibrium effective travel cost found is the minimum of effective route travel cost for each OD pair, and only those routes with their effective travel costs equal to the equilibrium effective travel costs carry flows, thereby satisfying the

reliability-based user equilibrium (RUE) conditions. The flow conservation condition is also satisfied, which can be verified from the results reported in Table 3.

One can observe that the complementarity condition for capacity constraints are satisfied in table 4: The passenger overload delay, d_s, is positive only when the corresponding link's residual capacity (defined as the difference between the effective link capacity and the corresponding link flow) is zero; when the residual capacity is positive, the passenger overload delay is zero. For links 7 and 8, although their overload delays equal zero, their residual capacities also equal zero because passengers on link 7 (8) competes line capacity with those on link 2 (4), and links 2 and 4 have been used up their effective capacities.

Not all the demands are met in this scenario. For OD pair 1, the virtual path carries flow and the equilibrium cost equals the predefined virtual path cost of $M = 1000$. It is therefore expected that there are critical links. In fact, the unsatisfied demand of OD pair 1 is due to insufficient bus service starting from JE, as links 1, 2, and 9 are all highly congested without residual capacities. Therefore, to increase the network capacity and the total met demand, the (critical) frequencies of bus service on these links 1, 2, and 9 have to be increased. For example, when the frequency of line 2 is increased from 7.7 to 12 veh/hr, the network capacity and the total met demand increase from 1209.9 to 1332.7 passengers/hr and the total unmet demand decreases from 790.1 to 667.3 passengers/hr.

When we determined the residual capacity of any path, which is the minimum of all the residual capacities of links on the path, we found that the path's residual capacity equals zero. This observation is consistent with the definition of the maximum network capacity that can be obtained only if all paths' residual capacities are used up.

Although the residual capacities of all routes including routes 5 and 10 equal zero, only the residual capacities of links 2 and 4 on R5 and R10 respectively equal zero (and their overload delays, d_s, are greater than zero). The residual capacities of links 3 and 5 are greater than zero because of the flow conservation condition and other critical links. For example, the residual capacity of 160.7 passengers/hr for link 5 cannot be utilized because the flow on link 5 must pass through either link 2 or link 4, but links 2 and 4 are carrying flow equal to their effective capacity.

The scenario also illustrates that the overload delay d_r^w on a route, the difference between effective travel cost and effective uncongested travel cost ($\eta_r^w - \tilde{\eta}_r^w$), does not equal to the sum of the overload delays d_s on the links on that route, even if a route only contains one link, because the route's overload delay considers the interaction between competing links. Taking route 2 (with only link 7) as an example, the overload delay d_s on link 7 is zero, but route 2's overload delay is 862.8 passengers/hr, which is the same as the overload delay d_s on link 2. The reason is that links 2 and 7 are competing and sharing the capacity of line 1. However, the total overload delay $\tilde{d}_s$ of link 7, which is the sum of the overload delays of links 2 and 7, is equal to the route 2's overload delay.

7.2 Effect of demand on equilibrium cost

To illustrate the effect of demand on equilibrium cost, Figure 3 is plotted. The demand for all OD pairs was varied from 50 passengers/hr to 250 passengers/hr. From this figure, it can be inferred that the equilibrium cost is stepwise increasing with demand in general. This is reasonable because the increase in demand may cause in-vehicle congestion and hence may cause the passengers to wait for the next arriving vehicle. Another reason is that the critical links belonging to the shorter routes are congested when demand increases, guiding

passengers to take a longer route, which results in delay in the shorter routes.

It should also be noted that the two graphs for OD pairs 1 and 2 show step functions but the other two graphs show horizontal lines. The reason is as follows: For OD pair 1, when the demand is low, the routes' overload delay is zero and does not influence the equilibrium cost. The equilibrium travel costs are only functions of the means and variances of in-vehicle travel and waiting costs which are not functions of demand. However, when the demand increases above a certain level, the routes' overload delay is positive and adds up the equilibrium cost. Hence, a step function can be seen with the jump occurring at certain demand level. The demand level at which the delay plays a role depends on the effective capacities of the links that serve the particular OD pair. For instance, consider OD pair 2, the equilibrium cost is 102.5 minutes until the demand is 136.5 passengers/hr. When the demand exceeds 136.5 passengers/hr, the cost increases to 105.5 minutes and then jumps to 107.3 minutes at the demand level of 241.5 passengers/hr. It is because when the demand does not exceed 136.5 passengers/hr, only R4(S9) carries flow and this path (link) has enough effective capacity to deal with the total demand of OD pair 2 and its competing link, link 2, also has enough capacity to cater the demand of OD pair 1. In order to fulfil the demand in excess of 136.5 passengers/hr, R3(S1) has to carry flow and the equilibrium cost of R4(S9) is increased from 102.5 to 105.5. This increment equals the routes' overload delay of 3.0 minutes on link 9 which operates at full effective capacity. Similarly, the second step increase is because the effective capacity of R3(S1) is reached, and the route's overload delay of 1.7 minutes is added to the equilibrium cost. In particular, the demands of OD pairs 1 and 2 at which the jumps occur are the same because the active routes in the two OD pairs contain the same competing and critical links (i.e., S2 and S9).

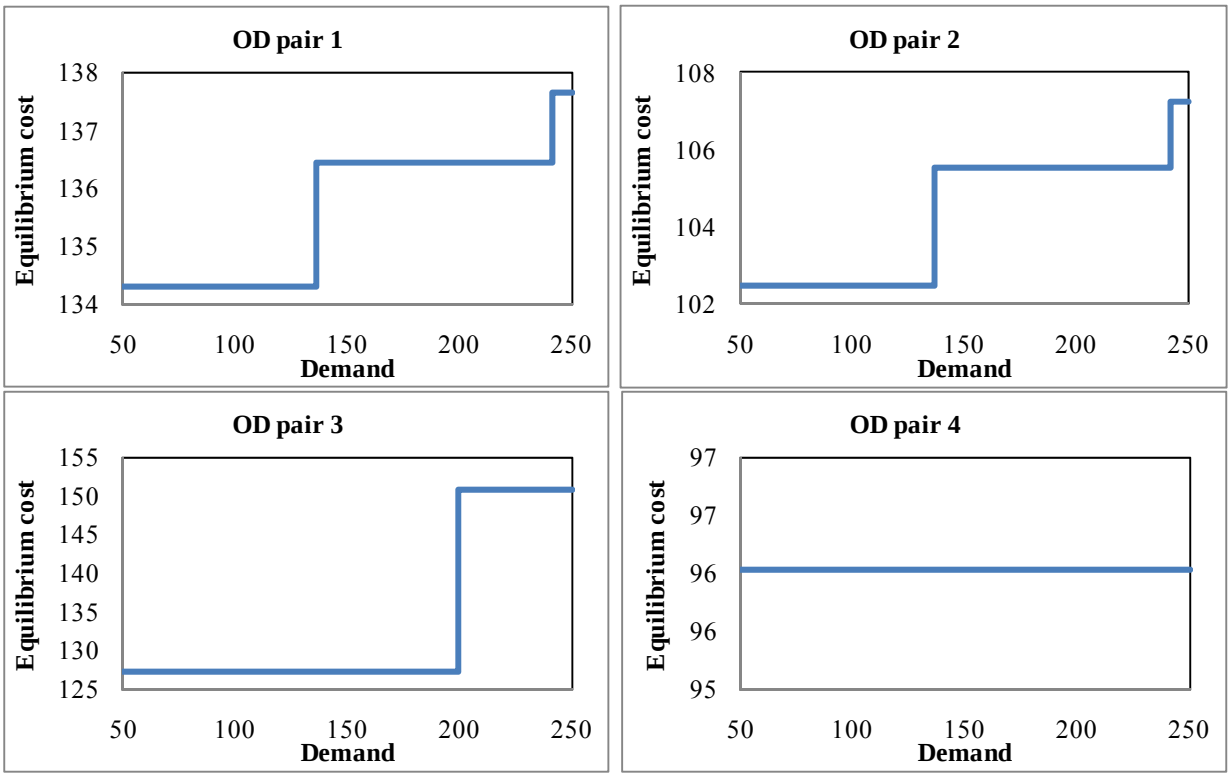

Figure 3 Equilibrium cost for various levels of demand

In the case of OD pair 4, the demand between each of these pairs is met without the routes' overload delay coming into play under the demand range considered. This means that the links connecting these OD pairs carry flows less than their effective path capacities.

Hence, the corresponding graph in Figure 5 is a straight line unlike the step function that was plotted for OD pairs 1, 2 and 3. If the demand range were wider, we could observe jumps as well. Moreover, if a more complicated and congested network with more paths between each OD pair were considered, more jumps on the equilibrium cost for each OD pair would be expected, where the number of jumps in each graph equals the number of used paths minus one, assuming all paths are not identical.

7.3 The effect of maximum violation probabilities and the degree of risk aversion of travellers on the flow pattern

This section illustrates the effect of maximum violation probabilities on flows and the degree of risk aversion of travellers on route choice. Figure 4 shows the residual capacities for various values of maximum violation probabilities when the demand of each OD pair is held at 250 passengers/hr. With an increase in the maximum violation probability of each link, we see that the available residual capacity increases for each link. From Eq. (24), we can infer that an increase in the maximum violation probability causes an increase in the effective capacity of each link, thereby increasing the residual capacity available on the link although the increase is not uniform across links. In particular, the residual capacity on S9 is 0 when α_s increases from 0.05 to 0.15, meaning that the increased effective capacity is fully utilized. This indicates that S9 is one of the critical links; hence, increasing the effective capacity of such link will increase the network capacity.

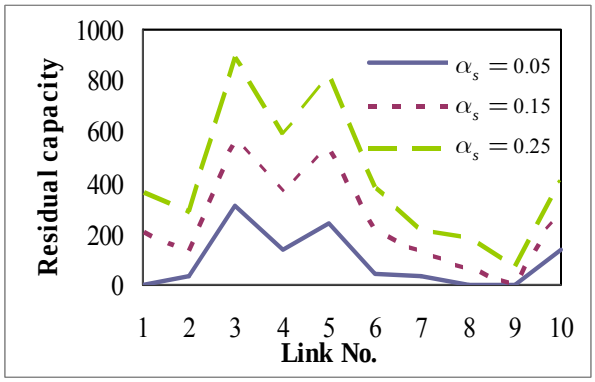

Figure 4 Residual capacities of links for various α_s values

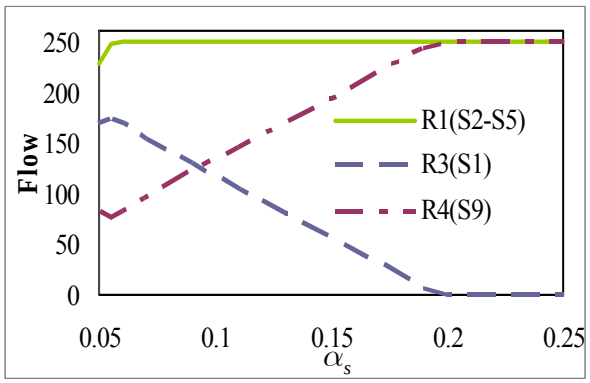

Figure 5 Influence of maximum violation probabilities on route choice

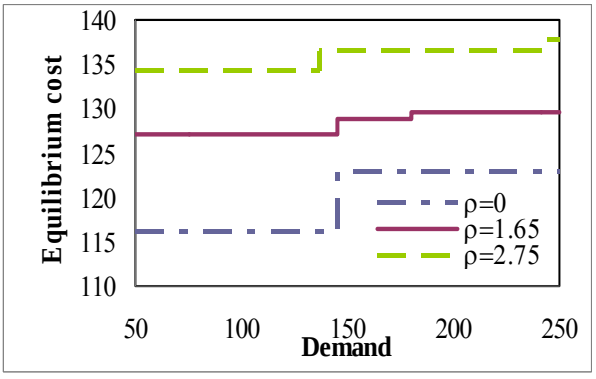

Figure 6 Influence of the degree of risk aversion of passengers between OD pair 1 on equilibrium cost

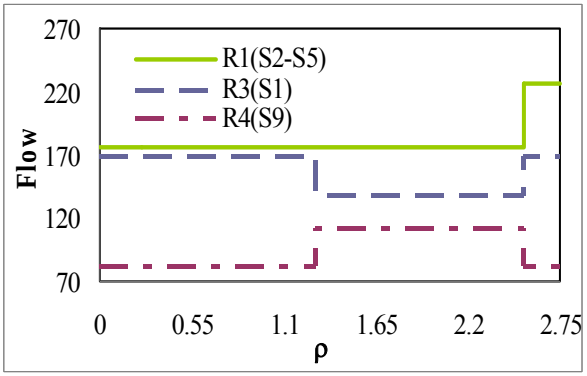

Figure 7 Influence of the degree of risk aversion of passengers on the flows on R1, R3 and R4

The flows on routes 1, 3 and 4 are plotted in figure 5, from which the influence of maximum violation probability, α_s, on route choice can be observed. In general, when the α_s value increases, more passengers can board R1 and R4 with the smallest effective uncongested travel cost because the effective capacities on these paths increase. In particular, the increase in α_s from 0.06 to 0.25 leads to a significant increase in flow on R4(S9), but the increase in α_s from 0.05 to 0.06 causes a small reduction of the flow on R4. This is because when α_s is larger than 0.06, the flow on R1 is stabilised to the level of the OD demand of 250, and the increase in the effective capacity of S2 will be totally used by passengers between OD pair JE-TP. However, when the flow on R1 has not been stabilised, the passengers on S2 will compete the effective capacity of line 2 with those on S9 and more passengers between OD pair JE-EU have a stronger motivation to use R1 (and hence S2) as the alternative route R2 between OD pair JE-EU is much worse than the alternative route R3 between OD pair JE-TP. Hence, the flow on R1 increases and the flow on R4 decreases.

The degree of risk aversion of passengers, as represented by ρ, influences the route choice pattern and the equilibrium travel cost of passengers. Figure 6 plots the equilibrium cost for passengers between OD pair 1 of different risk profiles under varying demands. From the figure, we see that the increase in the value of ρ also increases the equilibrium cost. This holds true for various levels of demand as well, and can be seen from Figure 6. This is because a more risk-aversive passenger has a larger safety margin and hence a larger equilibrium cost. Figure 7 shows the influence of ρ on route choice when the demand is fixed at 250 passenger/hr for each OD pair. When ρ is less than 1.29, only about 80 passengers are on R4. When ρ is between 1.29 and 2.53, more passengers select route 4 which has a lower path variance despite having a higher mean path cost. However, when ρ is larger than 2.53, the flow of R4 (S9) is reduced again due to the increase in flow on R1 (S2), which competes the capacity of line 2 with the flow on R4.

7.4 Benefit of considering variability in flow prediction

This sections aims at illustrating the potential benefit of using the proposed model. Two cases are considered: $\rho = 0$ (ignoring travel cost variability) and $\rho = 2.75$ (considering travel cost variability). For both cases, the demand for each OD pair is set to be 250 veh/hr and three performance indexes are calculated as shown in Table 5. The number in brackets under column $\rho = 0$ represents the percentage change with respect to the corresponding value under column $\rho = 2.75$. When the variability is not considered, total mean travel cost and effective travel cost for all the passengers are all underestimated because more the lowest mean cost routes with a higher variance are selected. More importantly, the total overload delay of all the passengers is highly overestimated when the variability is ignored. In fact, because of the variability, the risk aversive passengers will select routes with a higher probability of being punctual arrivals even if their mean costs are higher. This results in lower congestion on those lowest mean cost routes as reflected by a lower total overload delay. The implication is that the performance of transit networks could be evaluated improperly if the variability was ignored. From operators' point of view, ignoring the variability will lead to a wrong estimation on the flow pattern and a wrong allocation of resource to improve the transit network. By using the proposed model in which the variability is considered, the flow pattern and the congestion delay are estimated more accurately and hence the critical transit lines can be identified properly.

Table 5 Comparison between considering and ignoring variability

	$\rho = 2.75$	$\rho = 0$
Total mean travel cost	95111.3	94423.3 (-0.7%)
Total overload delay	6001.6	12665.8 (110.0%)
Total effective travel cost	122781.5	107089.1 (-12.8%)

7.5 Illustration of the solution method

The objective of this section is to illustrate the proposed solution method. We adopt the same network as in Figure 2, along with the associated data. The solution method is illustrated by solving the network with demand levels of 250, 250, 200 and 200 passengers/hr for OD pairs 1-4 respectively. Figure 8a shows the network flows and the objective value at the start of the solution method, in which every OD pair is connected by one virtual path (dotted lines) and is carrying flow equal to the OD demand. At each iteration, one path is generated by the K-shortest path algorithm and then added to the reduced problem (see figure 8b) that is solved by the revised simplex method. The optimal solution and the optimal basis of the current iteration are, respectively, used as the initial solution and the initial basis for the next iteration. This procedure is repeated (see figures 8c-d) until sufficient paths are added and no flows on all virtual paths can be observed as shown in Figure 8d.

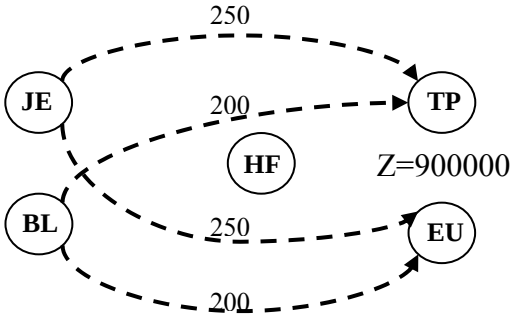

Figure 8a Initial solution

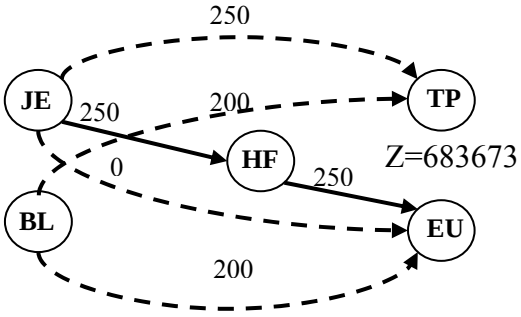

Figure 8b Solution after iteration 1

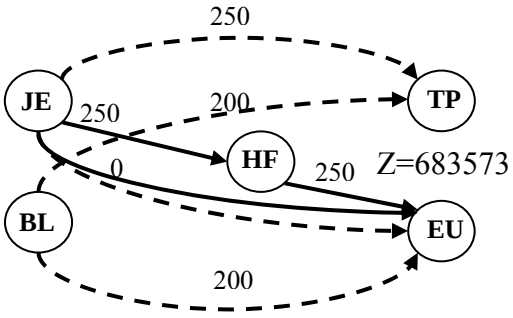

Figure 8c Solution after iteration 2

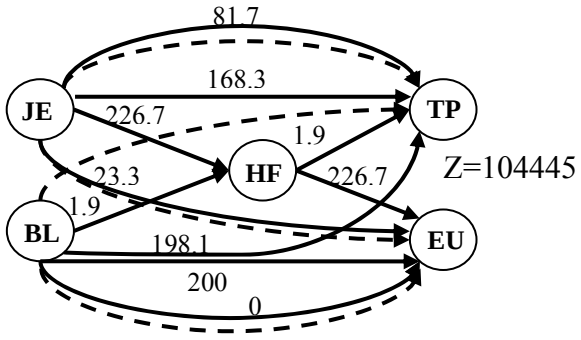

Figure 8d Final solution after iteration 10

Figure 8 Illustration of the proposed solution method.

In this example, we can observe the following:

1. The network is built incrementally.
2. The objective value is non-increasing with every iteration as shown in figure 9, which agrees with proposition 5 in the Appendix.
3. The objective value at iteration 1 is equal to that at iteration 2 and the objective values at iterations 4 and 5 are equal. It is because the new paths generated are not used. For

example, after iteration 2 (see Figure 10b), route (JE-EU) is generated, but it does not carry flow after solving the reduced problem due to its high cost.

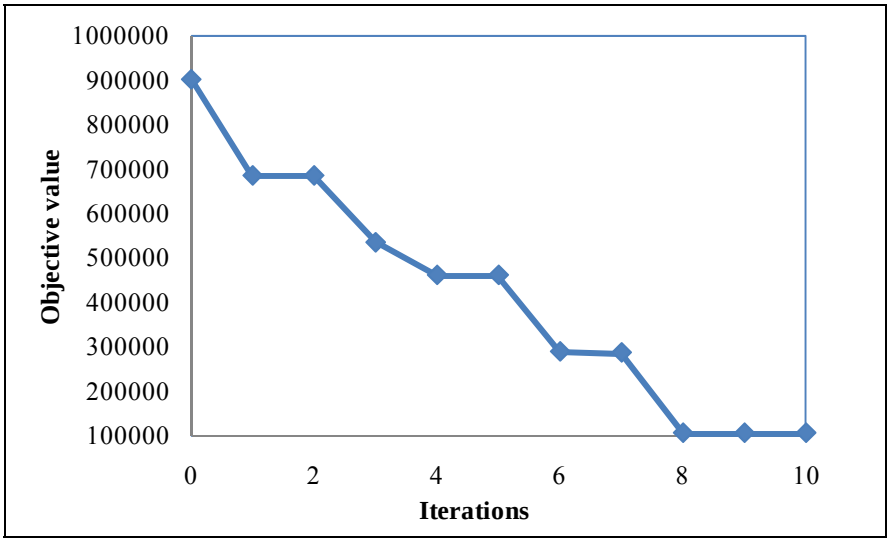

Figure 9 LP objective values over iterations

8. CONCLUSIONS

This paper proposes the reliability-based stochastic transit assignment problem that is formulated via the capacity constraint approach. The capacity constraint is defined based on the notion of effective capacity, and is developed via the chance constraint approach. Both in-vehicle travel times and waiting times are modelled as random variables and their means and variances are captured in the problem via the concepts of effective travel cost and reliability-based user equilibrium.

A LCP formulation is developed for the proposed problem and is later reformulated as a route-based linear programming problem (LP) with non-additive link costs. This route-based formulation allows us to determine the patronage, critical links, critical service frequencies, met and unmet demand, and the network capacity. This network capacity differs from the existing network capacity definition in the sense that the former captures the risk-averse behaviour of travellers. To solve this LP, a solution method is also proposed based on the column generation technique, the K-shortest path algorithm, and the revised simplex method. This proposed method guarantees finite convergence, and is illustrated by a simple example. Numerical studies are also set up to illustrate the properties of the proposed problem and the application of the proposed model for reliability analysis.

The proposed formulation has considered a constant capacity for all the transit vehicles in transit network. While this is not realistic, it is not difficult to extend the problem so that it can accommodate vehicles of different types and different capacities. The formulation proposed here incorporates only single-class passengers. As a topic of further interest, multiple user classes can be considered under the given setting. Moreover, the perception errors of passengers on in-vehicle travel time and waiting time have not been considered, thereby providing for an interesting avenue to explore further. Lastly, the assumption of exponential headway distribution is realistic to the transit stops without dynamic passenger information systems but may not be realistic to the stops with these systems. In the future, one can extend the proposed framework in this paper to consider the realistic assumption mentioned in Nökel and Wekeck (2009) for transit assignment under the provision of dynamic passenger information systems.

ACKNOWLEDGEMENT

The research was jointly supported by a grant (200902172003) from the Hui Oi Chow Trust Fund and two grants (201001159008 and 201011159026) from the University Research Committee of the University of Hong Kong.

REFERENCES

Abdel-Aty, M.A., Kitamura, R. and Jovanis, P.P. (1997) Using stated preference data for studying the effect of advanced traffic information on drivers' route choice. *Transportation Research Part C* 5, 39-50.

Ahuja, R.K., Magnanti, T.L., and Orlin, J.B. (1993) *Network Flows: Theory, Algorithms and Applications*. Prentice Hall, Englewood Cliffs, NJ.

Akamatsu, T. and Miyawaki, O. (1995) Maximum network capacity problem under the transportation equilibrium assignment (in Japanese). *Infrastructure Planning Review* 12, 719-729.

Asakura, Y. (1992) Maximum capacity of road network constrained by user equilibrium conditions. Paper presented at the 24th Annual Conference of the UTSG.

Bazaraa, M.S. and Jarvis, J.J. (1977) *Linear Programming and Network Flows*. John Wiley and Sons. New York.

Bouzaïene-Ayari, B., Gendreau, M. and Nguyen, S. (1995) An equilibrium-fixed point model for passenger assignment in congested transit networks. Technical Report CRT-95-57, Centre de Recherche sur les Transports, Univ. de Montreal.

Cepeda, M., Corninetti, R. and Florian, M. (2006) A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. *Transportation Research Part B* 40, 437-459.

Chen, A., Yang, H., Lo, H.K. and Tang, W. (1999) A capacity related reliability for transportation networks. *Journal of Advanced Transportation* 33, 183-200.

Chriqui, C. and Robillard, P. (1975) Common bus lines. *Transportation Science* 9, 115–121.

Cominetti, R. and Correa, J. (2001) Common-lines and passenger assignment in congested transit networks. *Transportation Science* 35, 250-267.

De Cea, J. and Fernández, E. (1989) Transit assignment to minimal routes: An efficient new algorithm. *Traffic Engineering and Control* 30, 491-494.

De Cea, J. and Fernández, E. (1993) Transit assignment for congested public transport systems: An equilibrium model. *Transportation Science* 27, 133-147.

Dial, R. B. (1967) Transit pathfinder algorithms. *Highway Research Record* 205, 67–85.

Fearnside, K. and Draper, D.P. (1971) Public transport assignment-a new approach. *Traffic Engineering Control* 12, 298-299.

Ge, Y.E., Zhang, H.M. and Lam, W.H.K. (2003) Network reserve capacity under the influence of traveler information. *Journal of Transportation Engineering,* 262–270.

Hamdouch, Y. and Lawphongpanich, S. (2008), Schedule-based transit assignment model with travel strategies and capacity constraints. *Transportation Research Part B,* 42, 663-684.

Jackson, W.B. and Jucker, J.V. (1982) An empirical study of travel time variability and travel choice behavior. *Transportation Science* 16, 460-475.

Kasikitwiwat, P. and Chen, A. (2005) Analysis of transportation network capacity related to different system capacity concepts. *Journal of the Eastern Asia Society for Transportation Studies* 6, 1439 - 1454

Kurauchi, F., Bell, M. and Schmöcker, J.-D. (2003) Capacity constrained transit assignment with common lines. *Journal of Mathematical Modelling and Algorithms* 2, 309-327.

Lam, W.H.K., Gao, Z.Y., Chan, K.S. and Yang, H. (1999) A stochastic user equilibrium assignment model for congested transit networks. *Transportation Research Part B* 33, 351-368.

Lam, W.H.K., Shao, H. and Sumalee, A. (2008) Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. *Transportation Research Part B* 42, 890-810.

Lam, W.H.K., Zhou, J. and Sheng, Z.H. (2002) A capacity restraint transit assignment with elastic line frequency. *Transportation Research Part B* 36, 919-938.

Le Clercq, F. (1972) A public transport assignment model. *Traffic Engineering Control* 13, 91-96.

Lee, D.H., Wu, L. and Meng, Q. (2006) Equity based land-use and transportation problem. *Journal of Advanced Transportation* 40, 75-93.

Leurent, F. (2011). On seat capacity in traffic assignment to a transit network. *Journal of Advanced Transportation*, in press.

Leurent, F. and Liu, K. (2009). On seat congestion, passenger comfort and route choice in urban transit: A network equilibrium assignment model with application to Paris. Paper presented at 88th Annual Transportation Research Board Meeting, Washington, DC, January 2009.

Li, Z.C., Lam, W.H.K. and Sumalee, A. (2008) Modeling impacts of transit operator fleet size under various market regimes with uncertainty in network. *Transportation Research Record, Journal of Transportation Research Board*, no. 2063, 18-27. National Research Council, Washington.

Li, Z.C., Lam, W.H.K. and Wong, S.C. (2009) The optimal transit fare structure under different market regimes with uncertainty in the network. *Networks and Spatial Economics* 9, 191-216.

Liu, J. and Deng, W. (2010) Network capacity research based on travel time reliability. *Proceedings of the 2010 International Conference on E-Product E-Service and E-Entertainment*, 1-3.

Lo, H.K., Luo, X.W. and Siu, B.W.Y. (2006) Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion. *Transportation Research Part B* 40, 792-806.

Morlok, E.K. and Riddle, S.P. (1999) Estimating the capacity of freight transportation systems: a model and its application in transport planning and logistics. *Transportation Research Record* 1653, 1–8.

Nguyen, S. and Pallottino, S. (1988) Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research* 37, 176-186.

Nguyen, S., Pallottino, S. and Malucelli, F. (2001) A modeling framework for passenger assignment on a transport network with timetables. *Transportation Science* 35, 238-249.

Nökel, K. and Wekeck, S. (2009) Boarding and alighting in frequency-based transit assignment. *Transportation Research Record* 2111, 60-67.

Poon, M.H., Tong, C.O. and Wong, S.C. (2003) Validation of a schedule-based capacity restraint transit assignment model for a large-scale network. *Journal of Advanced Transportation* 38, 5-26.

Poon, M.H., Tong, C.O. and Wong, S.C. (2004) A dynamic schedule-based model for congested transit networks. *Transportation Research Part B* 38, 343-368.

Schmöcker, J.-D., Bell, M.G.H. and Kurauchi, F. (2008) A quasi-dynamic capacity constrained frequency-based transit assignment model. *Transportation Research Part B* 42, 925-945.

Schmöcker, J.-D., Fonzone, A., Shimamoto, H., Kurauchi, F. and Bell, M.G.H. (2011) frequency-based transit assignment considering seat capacities. *Transportation Research*

Part B 45, 392-408.
Shao, H., Lam, W.H.K. and Tam, M. (2006) A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. *Networks and Spatial and Economics* 6, 173-204.
Siu, B.W.Y. and Lo, H.K., (2008) Doubly uncertain transportation network: Degradable capacity and stochastic demand. *European Journal of Operational Research* 191(1), 166-181.
Spiess, H. and Florian, M. (1989) Optimal strategies: A new assignment model for transit networks. *Transportation Research Part B* 23, 83-102.
Sumalee, A., Tan, Z.J. and Lam, W.H.K. (2009) Dynamic stochastic transit assignment with explicit seat allocation model. *Transportation Research Part B* 43, 895-912.
Sumalee, A., Uchida, K. and Lam, W.H.K. (2011) Stochastic multi-modal transport network under demand uncertainties and adverse weather condition. *Transportation Research Part C* 19, 338-350.
Szeto, W.Y. and Solayappan, M. (2009a) A variational inequality formulation for the reliability-based stochastic user equilibrium problem for congested transit networks. *Proceedings of the Eastern Asia Society for Transportation Studies* 7, 87-102.
Szeto, W.Y. and Solayappan, M. (2009b) A novel doubly stochastic transit assignment formulation: an application to Singapore bus network. *HKIE Transactions* 16, 63-71.
Szeto, W.Y., Jiang, Y, and Sumalee, A. (2012) A cell-based model for multi-class doubly stochastic dynamic traffic assignment. *Computer-Aided Civil and Infrastructure Engineering*, in press.
Szeto, W.Y., O'Brien, L. and O'Mahony, M. (2006) Risk-averse traffic assignment with elastic demands: NCP formulation and solution method for assessing performance reliability. *Network and Spatial Economics* 6, 313-332.
Szeto, W.Y., Solayappan, M. and Jiang, Y. (2011) Reliability-based transit assignment for congested stochastic transit networks. *Computer-Aided Civil and Infrastructure Engineering* 26, 311-326.
Teklu, F. (2008) A stochastic process approach for frequency-based transit assignment with strict capacity constraints. *Networks and Spatial Economics* 8, 225-240.
Tian, Q., Huang, H.J. and Yang, H. (2007a) Equilibrium properties of the morning peak-period commuting in a many-to-one mass transit system. *Transportation Research Part B* 41, 616-631.
Tian, Q., Huang, H.J. and Yang, H. (2007b) Commuting equilibria on a mass transit system with capacity constraint. In: Allsop, R.E., Bell, M.G.H., Heydecker, B.G. (Eds.), *Transportation and Traffic Theory* 2007. Elsevier, pp. 360-383.
Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K. (2007) *Probability & Statistics for Engineers and Scientists*. The 8th edition, Pearson Education International, NJ.
Wong, S.C. and Yang, H. (1997) Reserve capacity for a signal controlled road network. *Transportation Research Part B* 31, 397-402.
Wu, J., Florian, M. and Marcotte, P. (1994) Transit equilibrium assignment: A model and solution algorithms. *Transportation Science* 28, 193-203.
Yang, H., Bell, M.G.H. and Meng, Q. (2000) Modeling the capacity and level of service of urban transportation networks. *Transportation Research Part B* 34, 255-275.
Yang, L. and Lam, W.H.K. (2006) Probit-type reliability-based transit network assignment. *Transportation Research Record* 1977, 154-163.
Yin, Y., Lam, W.H.K. and Ieda, H. (2004) New technology and the modelling of risk taking behaviour in congested road networks. *Transportation Research Part C* 12, 171-192.
Yin, Y., Miller, M.A. and Lam, W.H.K. (2003) A simulation-based reliability assessment approach for congested transit network. *Journal of Advanced Transportation* 38, 27-44.

Ziyou, G. and Yifan, S. (2002) A reserve capacity model of optimal signal control with user-equilibrium route choice. *Transportation Research Part B* 36, 313-323.

APPENDIX

This appendix gives the proofs of Propositions 1-9.

Proposition 1: Under assumption 16, a solution exists to problem P(M).
Proof: Under assumption 16, the solution set is non-empty since one of the feasible solutions is that the virtual paths carry flows equal to the corresponding OD demands. Moreover, the solution set is bounded since each path flow must be nonnegative and cannot be greater than the corresponding OD demand. Furthermore, the solution set is closed because the set contains all the boundary points. As the set is both closed and bounded, the solution set must be compact. In addition, the objective function is linear and hence continuous on the solution set. Therefore, by Weierstrass' Theorem, a solution exists to problem *P(M)*. □

Proposition 2: Under assumption 16, multiple solutions may exist to problem P(M).
Proof: We prove this proposition by giving an example. Suppose there is an OD pair with two identical directed parallel paths. Assume the effective capacity of each path is larger than the demand d of this OD pair. Then, all optimal solutions of this problem can be described by $\mathbf{Y}^* = \left(y, d-y\right)^Y$, where $0 \le y \le d$.□

Proposition 3: Let $\mathbf{Y}_a^*$ and $\mathbf{Y}^*$ be the optimal solution of $\mathbf{Y}_a$ and $\mathbf{Y}$ in P(M), respectively. If $\mathbf{Y}_a^* = \mathbf{0}$, then $\mathbf{Y}^*$ is an optimal solution for problem P, and the optimal objective values for P and P(M) are both equal to $\mathbf{C}^T\mathbf{Y}^*$. If $\mathbf{Y}_a^* > \mathbf{0}$, there is no feasible solution for problem P.
Proof: Part of the proof here follows the analysis on pages 157-158 in Bazaraa and Jarvis (1977). Let $\mathbf{Y}$ be any feasible solution column vector for *P*. Then, $\left[\mathbf{Y}^T, \mathbf{0}^T\right]^T$ is feasible for *P(M)*. $\mathbf{Y}_a^* = \mathbf{0}$ implies that (1) $\mathbf{Y}^*$ must be feasible for *P*, and (2) $\left[\mathbf{Y}^{T*}, \mathbf{0}^T\right]^T$ is optimal for *P(M)* and $Z_M^* = \mathbf{C}^T\mathbf{Y}^* + 0 \le Z_p = \mathbf{C}^T\mathbf{Y} + 0$ by definition, which further implies $\mathbf{C}^T\mathbf{Y}^* \le \mathbf{C}^T\mathbf{Y}$. The above two conditions imply that $\mathbf{Y}^*$ must be optimal for *P*. By substituting $\left[\mathbf{Y}^{T*}, \mathbf{0}^T\right]^T$ and $\mathbf{Y}^*$ to Z_M and Z_p, respectively, we obtain $Z_M^* = Z_p^* = \mathbf{C}^T\mathbf{Y}^*$.

Suppose $\mathbf{Y}$ was feasible for problem *P*. Then $\left[\mathbf{Y}^T, \mathbf{0}^T\right]^T$ was feasible for *P(M)*. By definition, the optimal objective value of $\left[\mathbf{Y}^{T*}, \mathbf{Y}_a^{*T}\right]^T$ is not greater than that of $\left[\mathbf{Y}^T, \mathbf{0}^T\right]^T$. Then, $\mathbf{C}^T\mathbf{Y}^* + M\mathbf{Y}_a^* \le \mathbf{C}^T\mathbf{Y}$. Since $\mathbf{Y}_a^* > \mathbf{0}$ and *M* is a very large number, the left hand side of the inequality is very large. However, the right hand side is bounded by $\mathbf{C}^T\mathbf{Q}$, thereby rendering the inequality impossible. This implies that $\mathbf{Y}$ could not be a feasible solution and hence the second part of the proposition follows. □

Proposition 4: The maximum number of used paths in problem P(M) is $|\mathcal{S}| + |\mathcal{W}|$.
Proof: For any LP, only basic variables can have positive values. As there are $|\mathcal{S}| + |\mathcal{W}|$ constraints in Problems *P* and *P(M)*, there are $|\mathcal{S}| + |\mathcal{W}|$ basic variables and hence there are at

most $|\mathcal{S}|+|\mathcal{W}|$ variables with positive values. □

Proposition 5: The optimal objective value calculated by the algorithm is non-increasing with every iteration.

Proof: This is equivalent to saying $Z^*_{i+1} \le Z^*_i$. Let $\left[\mathbf{Y}_i^{*T}, \mathbf{Y}_a^{*T}\right]^T$ be optimal for $RP_i(M)$. Then, $\left[\left[\mathbf{Y}_i^{*T}, 0\right], \mathbf{Y}_a^{*T}\right]^T$ must be feasible for $RP_{i+1}(M)$. Since in minimization, the objective value of a feasible solution for $RP_{i+1}(M)$ cannot be smaller than Z^*_{i+1}, we have $Z^*_{i+1} \le \mathbf{C}_{i+1}{}^T\left[\mathbf{Y}_i^{*T}, 0\right]^T + M\mathbf{Y}_a^* = \left[\mathbf{C}_i{}^T, \hat{C}_{i+1}\right]\cdot\left[\mathbf{Y}_i^{*T}, 0\right]^T + M\mathbf{Y}_a^* = \mathbf{C}_i{}^T\mathbf{Y}_i^* + M\mathbf{Y}_a^* = Z^*_i$, where $\hat{C}_{i+1}$ is the effective travel cost on the new path added at iteration *i*+1 □

Proposition 6: The optimal objective value calculated by the algorithm at the next iteration must be lower than that at the current iteration if the new path added for the next iteration has a positive residual effective capacity before re-optimization, and the virtual path connecting the same OD pair as the new path carries flow before re-optimization.

Proof: W.L.O.G., let $F^x_{i+1} > 0$ be the residual capacity of the path added between OD pair *x* at the next iteration $i+1$ and let $\left[\mathbf{Y}_i^{*T}, \mathbf{Y}_a^{*T}\right]^T$ be optimal for $RP_i(M)$ with $\mathbf{Y}_a^* = \left[y_a^{*w}\right]$ and $y_a^{*x} > 0$. Moreover, let $\mathbf{Y}'_a = \left[y_a^{*w} - R_a^w\right]$ with $R_a^w = \min\left(y_a^{*w}, F^x_{i+1}\right)$ if $w = x$ and $R_a^w = 0$ otherwise. Then, $\left[\left[\mathbf{Y}_i^{*T}, 0\right], \mathbf{Y}_a^{*T}\right]^T$ must be feasible for $RP_{i+1}(M)$ and gives the objective value of $\mathbf{C}_i{}^T\mathbf{Y}_i^* + M\mathbf{Y}_a^* = Z^*_i$. This value is greater than the objective value $Z'_{i+1} = Z^*_i - R_a^w\left[M - \hat{C}_{i+1}\right]$ of another feasible solution $\left[\left[\mathbf{Y}_i^{*T}, R_a^w\right], \mathbf{Y}_a'^T\right]^T$ for $RP_{i+1}(M)$ as $M > \hat{C}_{i+1}$ and $R_a^w > 0$. Since by definition, $Z^*_{i+1} \le Z'_{i+1}$, we have $Z^*_{i+1} \le Z'_{i+1} < Z^*_i$. □

Proposition 7: Suppose a virtual path in the current reduced problem carries flow. Moreover, all binding links on the new path between the same OD pair as the virtual path added at the next iteration are part of another used route that is longer than the new path. Then, the optimal objective value at the next iteration obtained by the algorithm must be lower than that at the current iteration.

Proof: This proof is similar to that of proposition 6 except that we further define a path *p* generated at iteration $p < i+1$ with $\hat{C}_p > \hat{C}_{i+1}$ carrying a positive flow *y*, and that we further let the new route between OD pair *x* with cost $\hat{C}_{i+1}$ pass through all the binding links on path *p*. In addition, we need to set $R_a^w = \min\left(Y_a^{*w}, y\right)$ if $w = x$ and $R_a^w = 0$ otherwise. Then, we can show $Z^*_{i+1} \le Z'_{i+1} = Z^*_i - \min\left(Y_a^{*w}, y\right)\left[\hat{C}_p - \hat{C}_{i+1}\right] < Z^*_i$.

Proposition 8: The largest mean travel cost for each OD pair is non-decreasing with every iteration.

Proof: By definition, $\omega_i^w = \max\left\{E[\tilde{C}_r^w], \omega_{i-1}^w\right\} \Rightarrow \omega_i^w \ge \omega_{i-1}^w$. □

Proposition 9. When condition (i) is reached, the flows on the used paths in the current reduced problem are optimal for the corresponding used paths in problem P, the flows on

other paths in problem P are zeros, and the current objective value is optimal.

Proof. When condition (i) is reached, any new generated paths between an OD pair will have their effective uncongested travel costs greater than the current optimal effective travel cost for that OD pair since the safety margin is non-negative. Adding these longer paths to the reduced problem does not affect the objective value because the reallocation of flows to these longer paths and the virtual path cannot reduce the objective value. Hence, all these longer paths and the virtual paths for problem *P(M)* carry no flows at optimality, and the flows on the used paths in the current reduced problem are the optimal flows on the corresponding used paths in problem *P*. The conclusion then follows directly from Proposition 3. □