Reliability-based transit assignment for congested stochastic transit networks

By W.Y. SZETO[1], MUTHU SOLAYAPPAN, YU JIANG

Department of Civil Engineering

The University of Hong Kong

ABSTRACT

This paper proposes a Nonlinear Complementarity Problem (NCP) formulation for the risk-aversive stochastic transit assignment problem in which in-vehicle travel time, waiting time, capacity and the effect of congestion are considered as stochastic variables simultaneously and both their means and variances are incorporated into the formulation. A new congestion model is developed and captured in the proposed NCP formulation to account for different effects of on-board passengers and passengers waiting at stops. A reliability-based user equilibrium condition is also defined based on the proposed generalized concept of travel time budget referred to as effective travel cost, and is captured in the formulation. A column generation based algorithm is proposed to solve the NCP formulation. A survey was conducted to validate that the degree of risk aversion of transit passengers affects their route choices. Numerical studies were performed to demonstrate the problem and the effectiveness of the proposed algorithm. The results also show that underestimating the congestion effect and ignoring the risk aversion behavior can overestimate the patronage of transit service, which have profound implications on the profit of the operators involved and the development of transit network design models.

[1]Corresponding author. e-mail: drwszeto@yahoo.com.hk

1 INTRODUCTION

Transit assignment problems have received considerable attention over the past two decades. Some of the earliest work in the area of transit assignment can be traced to Dial (1967), Fearnside and Draper (1971) and Le Clercq (1972) in which the shortest path is computed after accounting for the waiting time at transit stops. However, the assumptions on fixed in-vehicle travel cost and expected travel time are very simplistic. Moreover, their models cannot deal with the route choice behavior of passengers at a transit stop shared by several competitive transit lines, often referred to as the common line problem.

Chriqui and Robillard (1975) are the first to deal with the common line problem by proposing the idea of the attractive set of transit lines between two consecutive stops as a subset of transit lines, which minimizes the passengers' expected travel time. The assignment of bus passengers was done proportionally to the nominal frequency of each common line. Following this, Spiess (1984) introduced the idea of strategy, which is a choice of an attractive set of lines at each boarding point. Later, using the idea of strategy, Nguyen and Pallottino (1988) presented a graph theoretic framework under the context of a hyperpath problem. Spiess and Florian (1989) proposed a linear programming formulation to determine the optimal strategy in a transit network. They assumed that the passengers will select a set of attractive lines and board the first arriving vehicle, thereby, minimizing the expected trip time.

Congestion related to overcrowded vehicles and stops is one of the key issues hampering the performance of transit systems in reality. This issue was also considered in parallel with the common line problem. For example, Nguyen and Pallottino (1988) considered the effect of congestion in the hyperpath model that they developed. Apart from introducing the concept of transit route and effective frequency, De Cea and Fernández (1993) also dealt with the effects of congestion at bus stops and aboard the transit vehicles. Cominetti and Correa (2001) investigated the network equilibrium model with congestion, in which

congestion affects both the waiting time and flow distribution. A queue-theoretic approach was adopted to model the congestion effects.

The concepts of deterministic user equilibrium (DUE) and stochastic user equilibrium (SUE) adopted in road networks have been introduced to transit assignment since late 1980's. The concept of DUE was first introduced to transit assignment by Nguyen and Pallottino (1988). Subsequently, many DUE transit assignment models were developed (e.g., Spiess and Florian, 1989; De Cea and Fernández, 1993; Cominetti and Correa, 2001; Cepeda *et al.*, 2006). However, these models assumed that the passengers have perfect knowledge about the network condition, which may not be realistic. Lam *et al.* (1999) utilized the idea of SUE to solve the transit assignment problem with capacity constraints in which passengers are assumed to select the lowest perceived travel cost routes. Lam *et al.* (2002) further proposed a SUE transit assignment model with congestion under the assumptions of the frequency on each transit line to be dependent on the vehicle dwelling time at each station and constant in-vehicle travel time. Lei and Chen (2004) also considered the SUE transit assignment with elastic demand and capacity constraint. They developed an algorithm based on the penalty function method to solve the problem.

The above DUE and SUE models were developed based on the approach of Chriqui and Robillard (1975), which is commonly referred to as the frequency-based approach. Although this approach ignores the detailed departure/arrival times, the frequency-based models are more computationally efficient and can handle larger transit networks. Such an approach is suitable for strategic and long term planning of large transit networks. However, according to Schmöcker *et al.* (2008), frequency-based approach cannot take into account the changing demand over time, the peak loading on transit vehicles and different levels of overcrowding at stations during the peak hours. Moreover, the departure time adjustments over days cannot be considered. Therefore, dynamic transit models (e.g., Poon *et al.*, 2004;

Schmöcker *et al.*, 2008; Teklu, 2008; Sumalee *et al.*, 2009) have been developed in the last decade.

Another aspect is that existing frequency-based models only consider mean waiting time and constant in-vehicle travel time but ignore the variabilities of the capacity and congestion. Moreover, these frequency-based models consider the mean trip time in determining the route choice of passengers and the influence of trip time variance in their route choice has not received much attention. Indeed, empirical studies like Abdel-Aty *et al.* (1997) and Jackson and Jucker (1982) pointed out that travel time variability plays a major role in influencing the trip makers' route choice behavior. Uncertain travel time causes trip makers including passengers to make a trade-off between travel cost and its uncertainty (Yin *et al.*, 2004). Such behavior is considered in traffic assignment (e.g., Bell and Cassir, 2002; Sumalee *et al.*, 2006) but to our best knowledge, this behavior has not received much attention in transit assignment. Moreover, in-vehicle travel time can be uncertain for buses and mini-buses as the in-vehicle travel time depends on both recurrent and non-recurrent congestion.

In a view to address these issues, we propose a stochastic approach to the frequency-based transit assignment problem that takes the variabilities of in-vehicle travel time, waiting time, congestion and capacity into account. These factors are modeled as random variables and both their means and variances are incorporated in the modeling framework so that both the network uncertainty and the risk aversion behavior of passengers can be captured. We define the reliability-based user equilibrium conditions based on the proposed generalized concept of travel time budget referred to as effective travel cost and formulate the transit assignment problem as a Nonlinear Complementarity Problem (NCP). A column generation based solution method is developed to solve the NCP formulation. Survey and numerical studies are carried out to validate the degree of risk aversion of transit

passengers affecting their route choices, and to demonstrate the problem and the effectiveness of the proposed solution method, respectively. The results also show that underestimating the congestion effect and ignoring the risk aversion behavior can overestimate the patronage of transit service, which have important implications on the profit of the operators involved and the development of transit network design models. Compared with the frequency-based transit assignment literature, the contributions of this paper include:

1) proposing a more realistic transit assignment formulation that
   - considers both demand and supply uncertainties,
   - captures risk-aversion behavior of passengers, and variabilities of in-vehicle travel time, waiting time, and congestion,
   - has at least one solution, and
   - can separately model different effects of on-board passengers and passengers waiting at stops on congestion cost.
2) developing an efficient solution method for the model that can apply to a realistic network, and
3) generalizing the concept of travel time budget to effective travel cost.

The rest of the paper is organized as follows. Section 2 describes the problem formulation. Section 3 depicts the solution method. Section 4 provides survey and numerical studies. Finally, Section 5 gives concluding remarks and identifies directions for future research.

2 PROBLEM FORMULATION

2.1 Network Representation, Definitions, and Assumptions

A transit network generally consists of a set of transit lines and stations (nodes) where passengers can board, alight or change vehicles. A transit line can be described by the

frequency of the vehicles (i.e. the number of vehicles of a transit line going across a screenline in a unit of time) and the vehicle types (e.g. bus or underground train). Note that in this paper the walk links will not be distinguished from the transit lines because it may be replaced by a transit line with a zero waiting time (very high service frequency). Different transit lines may run parallel for part of their itineraries with some stations in common. A line segment is a portion of any transit line between two consecutive stations of its itinerary and is characterized with a travel time and a frequency. A transit route is any path that a transit passenger can follow on the transit network in order to travel between any two nodes. Generally, it will be identified by a sequence of nodes, the first node being the origin of the trip, the final node being the destination and all the intermediate nodes being the transfer points. The portion of a route between two consecutive nodes is called route section, which is associated with a set of attractive lines or common lines. The set of attractive lines is assumed to be known and can be determined via the method in De Cea and Fernández (1993). Without loss of generality, a transit network can also be represented by a set of nodes and route sections.

For illustrative purposes, we adopt the network in De Cea and Fernández (1993) as an example. Figure 1 represents a transit network in terms of lines, while Figure 2 shows the same network coded by route sections. Table 1 illustrates the itinerary of the transit network in terms of transit routes, route sections and transfer nodes. In the example network shown, there is one origin-destination (OD) pair A-B, which is connected by four paths or routes. The four paths are formed by four different lines, each with different travel times and frequencies. For example, (25/10) on transit line L1 going from A to B, denotes a travel time of 25 minutes and a frequency of 10 buses/hour. We assume that a passenger waiting at a transfer node considers an attractive set of lines before boarding and knows the mean and variance of travel time of each line. The travel demand between each OD pair in the network is assumed to be

elastic. We also assume that a passenger selects the transit route that minimizes his/her effective travel cost discussed later as opposed to selecting the one that minimizes his/her travel time as in De Cea and Fernández (1993). Stochastic vehicle headways with the same distribution function are assumed for vehicles servicing different lines. However, the difference in vehicle headway traversing different lines could be achieved by varying the parameters of the distribution function.

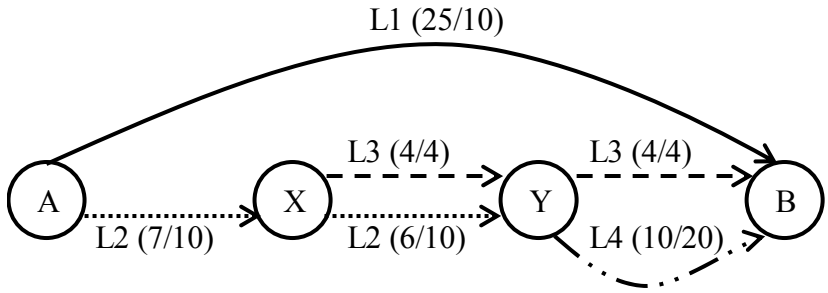


Figure 1: Transit network representation using transit lines

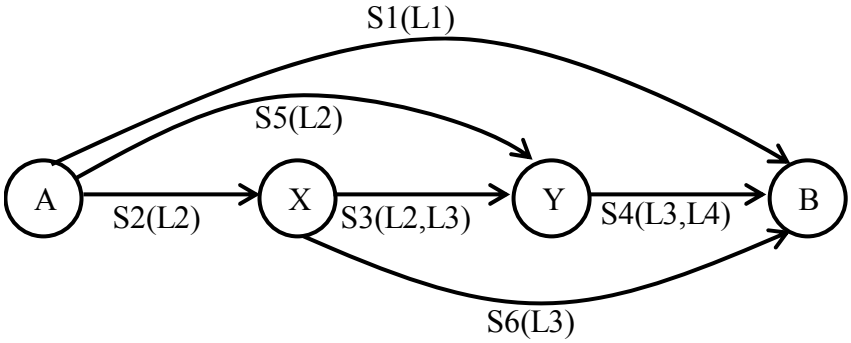


Figure 2: Transit network representation using route sections

Table 1: Transit routes and route sections

OD pairs	Transit Route	Route Sections (Transit Lines)	Nodes
A-B	1	S1(L1)	A,B
	2	S5(L2), S4(L3,L4)	A,Y,B
	3	S2(L2), S3(L2,L3), S4(L3,L4)	A,X,Y,B
	4	S2(L2), S6(L3)	A,X,B

In this paper, we consider a general transit network $G = (\mathcal{N}, \mathcal{S})$ when formulating the problem, where $\mathcal{N}$ refers to the set of nodes and $\mathcal{S}$ refers to the set of route sections

(links). The transit network has many paths and OD pairs. The set of OD pairs is denoted by $\mathcal{W}$ and the set of paths between OD pair $w \in \mathcal{W}$ is denoted by $\mathcal{R}_w$.

2.2 Effective Frequency

To model the effect of in-vehicle congestion in a transit network, we adopt a similar idea of effective frequency introduced by De Cea and Fernádez (1993). In a transit network constrained by its capacity, there is a positive probability that a transit vehicle arriving at a stop is full. Hence, passengers have to wait for the next arriving transit vehicle and this causes the frequency of the line at that particular stop to be effectively reduced from the passengers' point of view. This reduced line frequency is called effective frequency. In an ideal case, when there is no congestion, the effective frequency will be equal to its line frequency. Mathematically, the effective frequency can be expressed as:

$$f_s'^l = \frac{\alpha}{\frac{\alpha}{f^l} + \varphi_s^l}, \ \forall l \in A_s, \forall s \in \mathcal{S}, \tag{1}$$

where $f_s'^l$ is the effective frequency of line l on route section s. f^l is the frequency of line l. α is a positive parameter. φ_s^l is the additional waiting time for line l at stop $i(s)$, the origin node of route section s, due to in-vehicle congestion. A_s is the set of attractive lines associated with route section s.

The first term in the denominator in (1) is the waiting time under no in-vehicle congestion. α in this term is used to model the effect of different perceptions of waiting time and headway distributions (Spiess and Florian, 1989). When the unit of frequency is vehicles/hour and that of waiting time is minutes and when there is no perception error, the case $\alpha = 60$ min/hr corresponds to an exponential distribution of headways with mean $\frac{60}{f^l}$

minutes and the case $\alpha = 30$ min/hr approximates a constant headway of $\frac{60}{f^l}$ minutes. The first term plus the second term φ_s^l in the denominator in (1) gives the total waiting time under in-vehicle congestion. α divided by this sum gives the effective frequency. This derivation is in parallel to the case that frequency equals α divided by waiting time. One may notice that effective frequency depends on both route section and line, because the demand for service depends on both line and route section (or stop location). This contrasts to line frequency, which is the characteristic of a line, and is only line-specific.

In this paper, the additional waiting time for line l is expressed as:

$$\varphi_s^l = \beta^l \left(\frac{\bar{v}_{il}}{K^l} \right)^m, \forall l \in A_s, \forall s \in \mathcal{S}, \tag{2}$$

where K^l is the capacity of line l. $\bar{v}_{il}$ is the number of passengers per hour boarding line l before the origin node $i(s)$ of route section s and alighting after node $i(s)$. β^l and m are positive calibration parameters. The fraction $\frac{\bar{v}_{il}}{K^l}$ in (2) is interpreted as the occupancy rate, which is a measure of in-vehicle congestion. When the occupancy rate increases, the additional waiting time increases. Moreover, for a given occupancy rate, larger values of m and β^l mean that more passengers are willing to wait at the bus stop for the next arriving vehicle, leading to higher additional waiting time.

The capacity K^l of line l in Eq. (2) is given by:

$$K^l = f^l k, \ \forall l \in A_s, \forall s \in \mathcal{S}, \tag{3}$$

where k is the capacity of a transit vehicle and is assumed to be constant for all the vehicles servicing different routes for simplicity although there is no conceptual difficulty to generalize to the situation that different routes have different vehicle capacities.

2.3 Individual Cost Components

The cost on route section s, C_s, is described by three random variables:

$$C_s = \mu_T T_s + \mu_W \left(X_s + \varphi_s \right), \quad \forall s \in \mathcal{S}, \tag{4}$$

where T_s is the in-vehicle travel time on route section s. X_s is the waiting time for the first arrived vehicle on route section s that is not full. φ_s is the additional waiting time on route section s due to insufficient capacity. μ_T and μ_W are values of in-vehicle travel time and waiting time respectively. This section describes these individual cost components.

2.3.1 In-vehicle Travel Time

Let T_s^l, the in-vehicle travel time for line l on route section s, be a random variable. Then, the in-vehicle travel time on route section s can be found using the relation

$$T_s = \frac{\sum_{l \in A_s} f_s'^l T_s^l}{\sum_{l \in A_s} f_s'^l}, \quad \forall s \in \mathcal{S}. \tag{5}$$

Effectively, Eq. (5) calculates the weighted average of in-vehicle travel times. The expected in-vehicle travel time can be obtained by taking expectation on both sides of Eq. (5):

$$E[T_s] = \frac{\sum_{l \in A_s} f_s'^l E[T_s^l]}{\sum_{l \in A_s} f_s'^l}, \quad \forall s \in \mathcal{S}. \tag{6}$$

Assume the in-vehicle travel times of different lines are independent. The variance of in-vehicle travel time can then be found by:

$$Var[T_s] = \frac{\sum_{l \in A_s} \left(f_s'^l \right)^2 Var[T_s^l]}{\left(\sum_{l \in A_s} f_s'^l \right)^2}, \quad \forall s \in \mathcal{S}. \tag{7}$$

In practice, in-vehicle travel times between different line sections and between different lines sharing the same route section are not likely to be independent. When these in-vehicle travel times are highly correlated, covariance terms must be added to Eq. (7) to improve the accuracy of modeling.

2.3.2 Waiting Time for the First Arrived Vehicle

The waiting time distribution for the arrival of the first vehicle that is not full can be derived from the headway distribution of transit vehicles as discussed in Spiess and Florian (1989) but here we incorporate the concept of effective frequency in determining the mean and variance of waiting time for the first arrived vehicles. Assuming that passengers arrive at bus stops randomly, the waiting time distribution for line l on route section s can be determined by:

$$g_s^l(x) = \frac{[1 - H_s^l(x)]}{\int_0^\infty [1 - H_s^l(t)]dt}, \ \forall l \in A_s, \forall s \in \mathcal{S}, \tag{8}$$

where $H_s^l(x)$ is the cumulative distribution of the interarrival times (or headways) and the detailed derivation for this equation can be found in Larson and Odoni (1981) and Kulkarni (1995). By definition, the cumulative distribution function of waiting time for line l on route section s, denoted by $G_s^l(x)$, can then be obtained as:

$$G_s^l(x) = P\{X_s^l \le x\} = \int_0^x g_s^l(t)dt, \ \forall l \in A_s, \forall s \in \mathcal{S}, \tag{9}$$

where X_s^l is the waiting time for line $l \in A_s$.

Using Eqs. (8) and (9), we can determine the mean and variance of waiting time for a particular line l on route section s and those for route section s based on the assumed distribution of vehicle headway. While there are many distributions (e.g., triangular or uniform distributions) that can be assumed for vehicle headway to derive the analytical formula for the mean and variance, for the purpose of illustration, we assume the headway for

line l on route section s to be exponentially distributed with mean $\alpha/f_s'^l$. Hence, we have:

$$H_s^l(x) = 1 - e^{-\frac{f_s'^l}{\alpha}x}, \ \forall l \in A_s, \forall s \in \mathcal{S}. \tag{10}$$

Substituting Eq. (10) into Eq. (8) and then substituting the resulting expression into Eq. (9), we get:

$$G_s^l(x) = 1 - e^{-\frac{f_s'^l}{\alpha}x}, \ \forall l \in A_s, \forall s \in \mathcal{S}, \tag{11}$$

which means that the waiting time of line *l* on route section *s* is exponentially distributed with mean $\alpha/f_s'^l$.

The mean and variance of waiting time on route section *s* can be deduced from Eq. (11). For a positive value ϕ and a cumulative distribution function $F_{X_s}(x)$ for the waiting time X_s on route section s, the moments of X_s is given by:

$$E[X_s^{\phi}] = \phi \int_0^{\infty} t^{\phi-1} \{1 - F_{X_s}(t)\} dt, \ \forall s \in \mathcal{S}. \tag{12}$$

Based on Eq. (12), the first and second moments of X_s can be obtained as follows:

$$E[X_s] = \int_0^{\infty} \{1 - F_{X_s}(t)\} dt, \ \forall s \in \mathcal{S}, \text{ and} \tag{13}$$

$$E[X_s^2] = 2\int_0^{\infty} t\{1 - F_{X_s}(t)\} dt, \ \forall s \in \mathcal{S}. \tag{14}$$

Assuming that the waiting time on each line l of route section s to be independent of each other, the brace terms in Eqs. (13) and (14) can be expressed as:

$$1 - F_{X_s}(x) = P(X_s \geq x) = \prod_{l \in A_s} P(X_s^l \geq x) = \prod_{l \in A_s} \{1 - G_s^l(x)\}, \ \forall s \in \mathcal{S}. \tag{15}$$

Then, the first and second moments can be simplified by putting Eqs. (11) and (15) into both Eqs. (13) and (14) as follows:

$$E[X_s] = \int_0^{\infty} \prod_{l \in A_s} \{1 - G_s^l(t)\} dt = \frac{\alpha}{\sum_{l \in A_s} f_s'^l}, \ \forall s \in \mathcal{S}, \text{ and} \tag{16}$$

$$E[X_s^2] = 2\int_0^\infty t \prod_{l \in A_s} \{1 - G_s^l(t)\} dt = \frac{2\alpha^2}{(\sum_{l \in A_s} f_s'^l)^2}, \ \forall s \in \mathcal{S}. \tag{17}$$

Since the variance of X_s can be determined by:

$$Var[X_s] = E[X_s^2] - (E[X_s])^2, \ \forall s \in \mathcal{S}, \tag{18}$$

we can substitute Eqs. (16) and (17) into Eq. (18) to get:

$$Var(X_s) = \frac{\alpha^2}{(\sum_{l \in A_s} f_s'^l)^2}, \ \forall s \in \mathcal{S}. \tag{19}$$

2.3.3 Additional Waiting Time due to Congestion

The mean and variance of additional waiting time due to congestion (or congestion cost expressed as waiting time equivalent) are derived from the proposed congestion function, which is more general than the one proposed by De Cea and Fernández (1993). The congestion function for route section s is expressed as:

$$\varphi_s = \beta_s \left(\frac{aV^s + a\overline{V}_s + b\hat{V}_s}{K_s} \right)^n, \forall s \in \mathcal{S}, \tag{20}$$

where a, b, β_s, and n are calibration parameters. K_s is the capacity of route section s. V^s is the flow or number of passengers per hour on route section s. $\overline{V}_s$ is the total number of passengers per hour boarding at node $i(s)$ but the passengers will not transfer to another lines and finish their trips at the destination node of route section s. $\hat{V}_s$ is the number of passengers per hour boarding lines belonging to route section s before $i(s)$ and alighting after $i(s)$.

$\overline{V}_s + \hat{V}_s$ in Eq. (20) represents the passenger flow that compete with V^s for the capacities of the same set of lines. Unlike the congestion function proposed by De Cea and Fernádez, the proposed congestion function takes into account the number of passengers per

hour $V^s + \overline{V}_s$ waiting and boarding at stops, because some passengers may not be able to get into a bus due to too many passengers waiting at the stops. a and b in Eq. (20) are used to model different impacts of various flows to congestion cost (expressed as waiting time equivalent), as the congestion cost due to waiting at stops may be higher than that due to in-vehicle congestion. Normally, we set b to be equal to 1. The numerator in Eq. (20) can be interpreted as generalized occupancy.

The route section flow V_s in Eq. (20) can be obtained once all route flows on the route section are known:

$$V_s = \sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}_w} b_{sr} y_r^w, \quad \forall s \in \mathcal{S}, \tag{21}$$

where y_r^w is the flow on route r between OD pair w. b_{sr} is the route-section route incidence indicator, which equals 1 if route section s is a part of route r, and equals 0 otherwise.

The competing flows $\overline{V}_s$ and $\hat{V}_s$ in Eq. (20) are respectively calculated as follows:

$$\overline{V}_s = \sum_{l \in A_s} \sum_{r \in S_{ls}^+} v_l^r, \ \forall s \in \mathcal{S} \ , \text{and} \tag{22}$$

$$\hat{V}_s = \sum_{l \in A_s} \sum_{r \in \overline{S}_{ls}} v_l^r, \ \forall s \in \mathcal{S}, \tag{23}$$

where v_s^l is the number of passengers per hour on line l on route section s. S_{ls}^+ is the set of route sections going out from node $i(s)$ and containing line l but excludes route section s. $\overline{S}_{ls}$ is the set of route sections containing line l with their origin nodes before $i(s)$ and their destination nodes after $i(s)$. Assuming that the passengers board the first arrived transit vehicles, the line section flow v_s^l in Eqs. (22) and (23) can be found by:

$$v_s^l = \frac{f_s'^l}{\sum_{j \in A_s} f_s'^j} V_s, \quad \forall l \in A_s, \forall s \in \mathcal{S}. \tag{24}$$

The route section capacity K_s in Eq. (20) is defined as:

$$K_s = \frac{\gamma k}{h_s},\ \forall s \in \mathcal{S}, \tag{25}$$

where γ is a conversion factor, and h_s is the headway of transit vehicles on route section *s*. If the unit for headway is minutes and that for the capacity of a line is passengers per hour, then the conversion factor, γ = 60 min/hr.

Since headway is a random variable, the capacity is also a random variable according to Eq. (25) and hence the additional waiting time due to congestion is also a random variable according to Eq. (20). Substituting Eq. (25) into Eq. (20), and taking expectation and variance on both sides of the resulting expression, we get:

$$E\left[\varphi_s\right] = \beta_s \left(\frac{aV^s + a\overline{V}_s + b\hat{V}_s}{\gamma k}\right)^n E[(h_s)^n],\ \forall s \in \mathcal{S}, \text{ and} \tag{26}$$

$$Var\left[\varphi_s\right] = \beta_s^{\,2} \left(\frac{aV^s + a\overline{V}_s + b\hat{V}_s}{\gamma k}\right)^{2n} Var[(h_s)^n],\ \forall s \in \mathcal{S}, \text{ respectively.} \tag{27}$$

Since the headway for line l on route section s is exponentially distributed with mean $\alpha/f_s'^l$ (i.e., $h_s^l \sim Exp\left(\alpha/f_s'^l\right), \forall l, s$), according to the property of superposition of Poisson processes, $h_s \sim Exp\left(\alpha/f_s'\right), \forall s$, where

$$f_s' = \sum_{l \in A_s} f_s'^l,\ \forall s \in \mathcal{S}. \tag{28}$$

The expected value and variance of $(h_s)^n$ can then be found by:

$$E[(h_s)^n] = n\int_0^\infty t^{n-1} e^{-\frac{f_s'}{\alpha}t} dt = n!\left(\frac{\alpha}{f_s'}\right)^n,\ \forall s \in \mathcal{S}, \text{ and} \tag{29}$$

$$Var[(h_s)^n] = E[(h_s)^{2n}] - (E[(h_s)^n])^2, \ \forall s \in \mathcal{S}$$
$$= 2n\int_0^\infty t^{2n-1} e^{-\frac{f_s'}{\alpha}t} dt - \left(n!\left(\frac{\alpha}{f_s'} \right)^n \right)^2, \ \forall s \in \mathcal{S} \quad (30)$$
$$= \left((2n)! - (n!)^2 \right)\left(\frac{\alpha}{f_s'} \right)^{2n}, \ \forall s \in \mathcal{S}.$$

Substituting Eqs. (29) and (30) into Eqs. (26) and (27) respectively, we obtain the expected value and variance of the additional waiting time due to congestion on route section s as shown below:

$$E\left[\varphi_s\right] = \beta_s n!\left(\frac{\alpha(aV^s + a\overline{V}_s + b\hat{V}_s)}{\gamma k f_s'} \right)^n, \ \forall s \in \mathcal{S}, \text{ and} \quad (31)$$

$$Var\left[\varphi_s\right] = \beta_s^2((2n)! - (n!)^2)\left(\frac{\alpha(aV^s + a\overline{V}_s + b\hat{V}_s)}{\gamma k f_s'} \right)^{2n}, \ \forall s \in \mathcal{S}. \quad (32)$$

2.4 Effective Travel Cost

The variabilities associated with the in-vehicle travel time and waiting time, coupled with the effect of congestion cause variability in route travel time. Due to this, passengers cannot determine the exact trip time for their journeys. The variability in route travel time is countered by early departures to allow for additional time to avoid late arrivals. This additional time is included by the passengers while planning their trips, and is referred to as travel time margin. This travel time margin plus the expected trip time is known as travel time budget (Lo *et al.,* 2006). However, this concept does not consider the fact that the monetary value of in-vehicle travel time differs from that of the waiting time. Hence, this paper proposes the concept of effective travel cost (or travel cost budget), which generalizes the concept of travel time budget by considering trip travel cost (including in-vehicle travel time cost and waiting time cost) instead of trip travel time. Mathematically, the effective travel cost on a particular route can be formulated as:

$$E_r^w = E[C_r^w] + M_r, \forall r \in \mathcal{R}_w, w \in \mathcal{W}, \tag{33}$$

where E_r^w is the effective travel cost of route r between OD pair w. C_r^w is the trip travel cost on route r connecting OD pair w. M_r is the travel cost margin of passengers using route r.

The travel cost margin M_r is expressed in terms of the standard deviation of trip travel cost:

$$M_r = \rho\sqrt{Var(C_r^w)}, \forall r \in \mathcal{R}_w, w \in \mathcal{W}, \tag{34}$$

where ρ is a parameter.. Similar to Lo *et al.* (2006), the parameter ρ can formally relate to the probability λ that the actual trip travel cost is not greater than effective travel cost:

$$P\{C_r^w \le E_r^w = E[C_r^w] + \rho\sqrt{Var(C_r^w)}\} = \lambda, \tag{35}$$

By Central limit theorem, a probability distribution tends to be a normal distribution when the sample size is large enough. Hence, it is reasonable to assume that C_r^w is normally distributed, and the random variable C_r^w can be normalized as shown below:

$$P\left\{\frac{C_r^w - E[C_r^w]}{\sqrt{Var(C_r^w)}} \le \rho\right\} = \lambda. \tag{36}$$

Let $Z_{C_r^w} = \dfrac{C_r^w - E[C_r^w]}{\sqrt{Var(C_r^w)}}$ denote the standard normal variate of C_r^w and hence Eq. (36) can be written as:

$$P(Z_{C_r^w} \le \rho) = \lambda. \tag{37}$$

The parameter ρ in (34) and hence λ in (37) can then be interpreted as the degree of risk aversion of passengers. A higher value of ρ means that the passenger is more risk-aversive and is willing to have a higher probability of the trip travel cost not greater than the effective travel cost. Thus, the values of ρ and λ totally depend on the individual's appetite for

risk aversion. These values also depend on the purpose of the trip. A more important trip will lead to a higher λ value and hence a higher ρ value.

The route costs in Eqs. (33) and (34) are related to route section costs as follows:

$$C_r^w = \sum_{s \in S} b_{sr} C_s, \ \forall r \in \mathcal{R}_w, \forall w \in \mathcal{W}. \tag{38}$$

Assume the variances of route costs are independent of each other. We can take expectation and variance on both sides of Eq. (38) to get the following respectively:

$$E[C_r^w] = \sum_{s \in S} b_{sr} E[C_s], \ \forall r \in \mathcal{R}_w, \forall w \in \mathcal{W}, \text{ and} \tag{39}$$

$$Var[C_r^w] = \sum_{s \in S} (b_{sr})^2 Var[C_s] = \sum_{s \in S} b_{sr} Var[C_s], \ \forall r \in \mathcal{R}_w, \forall w \in \mathcal{W}. \tag{40}$$

The effective travel cost on route *r* between OD pair *w* can then be obtained by substituting Eqs. (34), (39), and (40) into Eq. (33):

$$E_r^w = \sum_{s \in S} b_{sr} E[C_s] + \rho \sqrt{\sum_{s \in S} b_{sr} Var[C_s]}, \ \forall r \in \mathcal{R}_w, \forall w \in \mathcal{W}. \tag{41}$$

The mean and variance of route section cost in (41) can be found by:

$$E[C_s] = \mu_T E[T_s] + \mu_W E[X_s] + \mu_W E[\varphi_s], \text{ and} \tag{42}$$

$$Var[C_s] = \mu_T^2 Var[T_s] + \mu_W^2 Var[X_s] + \mu_W^2 Var[\varphi_s], \tag{43}$$

which are obtained by taking expectation and variance on both sides of Eq. (4) respectively. The effective route travel cost E_r^w can then be expressed in terms of individual components of route section costs:

$$\begin{aligned} E_r^w = & \sum_{s \in S} b_{sr} \left(\mu_T E[T_s] + \mu_W E[X_s] + \mu_W E[\varphi_s] \right) \\ & + \rho \sqrt{\sum_{s \in S} b_{sr} \left(\mu_T^2 Var[T_s] + \mu_W^2 Var[X_s] + \mu_W^2 Var[\varphi_s] \right)}, \ \forall r \in \mathcal{R}_w, \forall w \in \mathcal{W}. \end{aligned} \tag{44}$$

2.5 Nonlinear Complementarity Problem Formulation

Assuming that all the passengers choose the routes with minimum effective travel cost, we

define the reliability-based user equilibrium as follows:

Reliability-based user equilibrium: *The transit network is said to be at reliability-based user equilibrium, if, for each OD pair, the effective travel costs on used routes are equal to each other and are not greater than those on unused routes.*

The reliability-based user equilibrium as defined above can mathematically be stated as follows:

$$E_r^w \begin{cases} = & u_w, \quad y_r^w > 0, \\ \geq & u_w, \quad y_r^w = 0, \end{cases} \forall r \in \mathcal{R}_w, \forall w \in \mathcal{W}, \tag{45}$$

where u_w is the equilibrium effective travel cost over all the routes that connect OD pair $w \in \mathcal{W}$ and y_r^w is the passenger flow on route $r \in \mathcal{R}_w, w \in \mathcal{W}$. The nonlinear complementarity conditions for the routes on the network, based on those in Eq. (45) can be stated as follows:

$$E_r^w - u_w \geq 0, \ \forall r \in \mathcal{R}_w, \forall w \in \mathcal{W}, \ \text{and} \tag{46}$$

$$y_r^w (E_r^w - u_w) = 0, \forall r \in \mathcal{R}_w, \forall w \in \mathcal{W}. \tag{47}$$

Apart from the nonlinear complementarity conditions, the following flow conservation and non-negativity constraints for route flows also form a part of the formulation:

$$\sum_{r \in \mathcal{R}_w} y_r^w = q_w, \quad \forall w \in \mathcal{W}, \tag{48}$$

$$y_r^w \geq 0, \ \forall r \in \mathcal{R}_w, \forall w \in \mathcal{W}. \tag{49}$$

In this study, the demand in (48) is assumed to be elastic, and the following linear decreasing function is adopted for the purpose of analysis:

$$q_w = q_0 - \chi_w u_w, \quad \forall w \in \mathcal{W}, \tag{50}$$

where q_0 is the maximum or potential demand and χ_w is the slope of the demand function of OD pair $w \in \mathcal{W}$.

The reliability-based transit assignment problem can be formulated as a Nonlinear

Complementarity Problem (NCP): to find $\mathbf{Y} = [y_r^w] \geq \mathbf{0}$ such that

$$\mathbf{F}(\mathbf{Y}) \geq \mathbf{0}, \ \mathbf{Y}^T \mathbf{F}(\mathbf{Y}) = 0, \tag{51}$$

where the mapping function $\mathbf{F}(\mathbf{Y}) = \left[E_r^w - \frac{q_0 - \sum_{r \in \mathcal{R}_w} y_r^w}{\chi_w} \right]$. E_r^w is defined by (1)-(3), (6), (7), (16), (19)-(24), (28), (31), (32), and (44).

The NCP (51) can be reformulated as a variational inequality (VI) problem (see Nagurney, 1999): to find $\mathbf{Y} = [y_r^w]$ such that

$$\mathbf{F}^T(\mathbf{Y}^*)(\mathbf{Y} - \mathbf{Y}^*) \geq 0, \forall \mathbf{Y} \subseteq \Omega. \tag{52}$$

where Ω is the solution set. The superscript * refers to $\mathbf{Y}$ that satisfies (45). According to Nagurney (1999), a solution exists to (52) when $\mathbf{F}^T(\mathbf{Y})$ is continuous with respect to $\mathbf{Y}$ and the solution set is bounded and closed (i.e., compact). In addition, the solution is unique when $\mathbf{F}^T(\mathbf{Y})$ is strictly monotonic with respect to $\mathbf{Y}$. Clearly, the solution set is compact in this problem. It is because the flow cannot be greater than demand and hence the solution set must be bounded by a sphere with radius equal to the largest demand of all OD pairs. Moreover, $\mathbf{F}^T(\mathbf{Y})$ is continuous with respect to $\mathbf{Y}$ as all the functions involved for calculating E_r^w and $\frac{q_0 - \sum_{r \in \mathcal{R}_w} y_r^w}{\chi_w}$ are continuous of y_r^w. Therefore, a solution exists to this VI problem and hence to the NCP formulation as well. However, the monotonic requirement in this problem may not be satisfied, leading to the possibility of having multiple route flow solutions.

3 SOLUTION METHOD

The NCP formulation is path-based, which cannot be reformulated into a link-based

formulation because the standard deviation of the travel time on a path is path-specific and not equal to the sum of the standard deviations of the link travel times on that path. Therefore, we need to develop solution methods to handle the path-based formulation directly. For realistic networks, there are many paths and hence the path set and the number of path flow variables are very large. Enumerating all the paths in advance is very time-consuming but not all paths will be used at optimality. Moreover, handling too many variables will increase the computation time and computer storage may be a problem. Therefore, we develop a path generation algorithm to avoid the computationally intensive path enumeration and develop a data structure to store path flow variables with non-zero values only.

The proposed path generation algorithm is based on the algorithm in Chen *et al.* (2001) which is used to solve traffic equilibrium problem with path-specific tolls. A subroutine is developed to identify the lowest effective travel cost path in each major iteration. This subroutine uses a *k*-shortest path algorithm to find *k* lowest mean travel cost paths. Their travel cost variances and effective travel costs are then computed and the lowest effective travel cost path for each OD pair is identified. This path will then be added to the path set if the path has not been included yet. The algorithm also utilizes the self-adaptive projection and contraction algorithm proposed by Chen *et al.* (2001) to solve the NCP with the updated path set.

There are two main differences between the proposed algorithm and the one proposed by Chen *et al.* (2001). First, column dropping is not used to ensure the convergence of the algorithm under the general monotone mapping assumption for $\mathbf{F}^T(\mathbf{Y})$. Second, the path specific travel cost margins are functions of route flows but path specific tolls are not. Therefore, one subroutine is required to calculate the path-specific travel cost margins and effective travel cost in each iteration.

The detailed algorithmic steps are described as follows:

Step 1. Initialization

- Initialize parameters: terminating threshold $\kappa > 0$.
- Set main iteration counter $m = 0$.
- Perform incremental assignment to generate an initial set of paths: $\mathcal{R}_w(m), \forall w \in \mathcal{W}$, where $\mathcal{R}_w(m)$ is the path set of OD pair *w* in iteration *m*.
- Set flows on initial set of paths to be zero.

Step 2. Column Generation

- Increase main iteration counter: $m = m + 1$.
- Update the mean and variance of route section costs.
- Identify the lowest effective travel cost paths:
  - determine the *k* – lowest mean route travel cost paths.
  - calculate the variances of effective travel costs and then the path-specific travel cost margins of all the *k* – lowest mean route travel cost paths.
  - obtain the effective path travel costs on all the *k* – lowest mean route travel cost paths.
  - identify the path $\bar{p}^w(m)$ with the lowest effective travel cost.

Step 3. Convergence

- If the NCP's error bound $G = \max_{\forall r \in \mathcal{R}_w(m-1), \forall w \in \mathcal{W}} \left[y_r^w - \max\left(0, y_r^w - \frac{q_0 - \sum_{r \in \mathcal{R}_w} y_r^w}{\chi_w}\right)\right] \le \kappa$, then terminate. Otherwise, update the path set: $\mathcal{R}_w(m) = \bar{p}^w(m) \cup \mathcal{R}_w(m-1)$ if $\bar{p}^w(m) \notin \mathcal{R}_w(m-1), \forall w \in \mathcal{W}$, and go to Step 4.

Step 4. Equilibration

- Use the self-adaptive projection and contraction algorithm to solve the NCP using the path set $\mathcal{R}_w(m), \forall w \in \mathcal{W}$.
- Return to Step 2.

The convergence of this solution method depends on whether the self-adaptive projection and contraction algorithm can solve the NCP in each iteration, because in the worst case, all paths are included in the path set and the solution method becomes handling the original NCP. If $\mathbf{F}^T(\mathbf{Y})$ is monotone, the proposed solution method can guarantee convergence.

4 SURVEY AND NUMERICAL STUDIES

Three studies were carried out. The first one is to validate that the degree of risk aversion of transit passengers affects their route choices. The second one is to illustrate the properties of the problem and the last one is to illustrate the effectiveness of the algorithm.

Example 1: Validation on transit route choice behavior

To validate that the degree of risk aversion of passengers affects transit route choice, we conducted a survey with a sample size of 50 people in June 2009 in Singapore. Other than collecting some basic information on the respondents, the survey asked the respondents their choices from two given alternatives in each of the 6 different scenarios. Alterative 1 gives a usual trip time of 30 minutes in all scenarios whereas alternative 2 gives a lower usual trip time of 20 minutes but the possible delay increases from scenarios (a) to (f). This survey setting is similar to the one in Jackson and Jucker (1982) except that this survey focuses only on the transit services given. The details of the second part of the questionnaire are given in the Appendix, and the results are reported in Figure 3a.

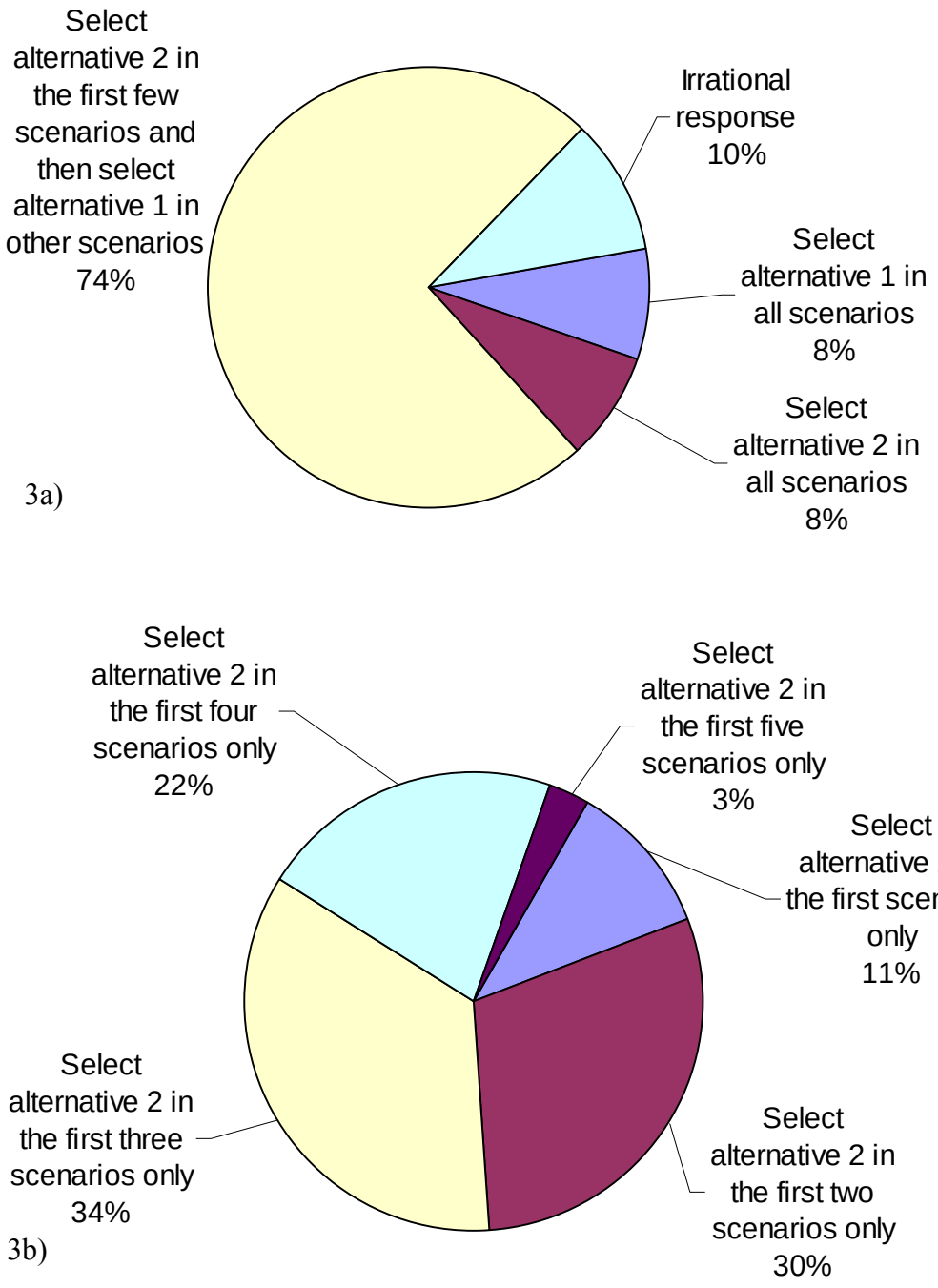


Figure 3: Survey result

10% of the respondents gave irrational answers and their answers were ignored. From the remaining respondents, we find that all the respondents have different degrees of risk aversion. 8% of them are risk-aversive and always select alternative 1. 8% of them are risk-neutral and always select the route with lower mean travel time regardless of delay or variance of travel time. All other respondents selected routes by making a tradeoff between

the mean and variance of route travel time. These 74% respondents are further divided into 5 classes as shown in Figure 3b. From this figure, most of the people select alternative 2 in the first two scenarios. The ρ value for these people is between 1.17 and 2, and is estimated based on the method in Jackson and Jucker (1982). Overall, the survey shows that the degree of risk-aversion highly affects the transit route choice. The implication is that ignoring this risk averseness in transit assignment can wrongly estimate the transit flow pattern and hence the level of service of each transit line.

Example 2: Properties of the problem

Table 2: Travel time and variance for line segments

t_s^l	t_1^1	t_2^2	t_3^2	t_3^3	t_4^3	t_4^4	t_5^2	t_6^3
Travel Time (min)	25	7	6	4	4	10	13	8
Variance (min^2)	3	12	12	8	18	22	35	14

To illustrate the properties of the problem clearly, we adopt the small network shown in Figures 1 and 2. The basic route section data related to the network is given in Table 2, which is similar to the one presented in De Cea and Fernández (1993). All transit lines are assumed to be served by the single-deck bus, Mercedes Benz O 405, which is currently operated in Singapore to serve the entire network. This particular bus has a seating capacity of 47 passengers and standing capacity of 38 passengers. Hence, the total capacity of the transit vehicle is 85 passengers. The headway is assumed to be exponentially distributed with mean $1/f_s^l$. The value of in-vehicle travel time is SGD 18.27 per hour, which is estimated by the average monthly salary of SGD 3977 for the year 2008 (Ministry of Manpower, Singapore, 2009a) and the average weekly paid working hour rate (including overtime) of SGD 50.1 for the same year (Ministry of Manpower, Singapore, 2009b). The value of waiting time is set to be twice as that of in-vehicle time, based on the two values suggested by the US Department

of Transportation and the recommendation by the UK Department of Transport (Victoria Transport Policy Institute, 2009). We also set $m = n = 4$, $\beta^l = 1$, $\beta_s = 0.1$, $a = b = 1$, $\alpha = \gamma = 60$ min/hr, $\lambda = 95\%$, and $\chi_w = 1$ unless otherwise specified.

The effects of potential demand, degree of risk aversion and congestion parameter value on route flows and route choice were studied using Table 3, which shows the results obtained for four different cases using the proposed solution method with $\kappa = 10^{-3}$ and $k = 5$. Case 1 is the base case and others differ from the base case by having one different parameter value. From Table 3, we can see that the reliability-based user equilibrium conditions are satisfied in all cases – all used routes have equal and minimal effective travel cost. However, these routes have different means and variances in each of the cost components.

Table 3: Route flows and route costs under different degrees of risk aversion, congestion parameter values and potential demand

Case	Path	Path Flow	Effective Travel Cost	In-vehicle Travel Time (min)		Waiting Time (min)		Additional Waiting Time due to Congestion (min)	
				Mean	Variance	Mean	Variance	Mean	Variance
1: n = 3, λ = 99%, $q_0 = 2000$	1	1089.4	23.6	25.0	3.0	6.0	36.0	1.3	30.3
	2	886.9	23.6	22.0	50.8	8.5	42.3	0.7	8.9
	3	0.0	28.4	21.4	34.1	13.4	65.9	1.1	11.4
	4	0.0	41.3	15.0	26.0	21.0	261.0	0.7	8.8
2: n = 3, λ = 99%, $q_0 =$ **400**	1	380.1	19.9	25.0	3.0	6.0	36.0	0.1	0.1
	2	0.0	22.4	22.0	50.8	8.5	42.3	0.0	0.0
	3	0.0	26.1	21.4	34.1	12.8	60.6	0.0	0.0
	4	0.0	40.5	15.0	26.0	21.0	261.0	0.0	0.0
3: n = **1,** λ = 99%, $q_0 = 2000$	1	1980.0	20.0	25.0	3.0	6.0	36.0	0.2	0.1
	2	0.0	22.4	22.0	50.8	8.5	42.3	0.0	0.0
	3	0.0	26.1	21.4	34.1	12.8	60.6	0.0	0.0
	4	0.0	40.5	15.0	26.0	21.0	261.0	0.0	0.0
4: n = 3, λ = **50%**, $q_0 = 2000$	1	1171.3	12.2	25.0	3.0	6.0	36.0	1.6	46.8
	2	816.4	12.2	22.0	50.8	8.5	42.3	0.6	5.4
	3	0.0	15.1	21.4	34.1	13.2	64.4	0.8	6.6
	4	0.0	17.7	15.0	26.0	21.0	261.0	0.5	5.4

By comparing cases 1 and 2 in Table 3, we can observe that equilibrium effective

travel cost increases with increasing potential demand. Moreover, similar patterns can be observed for the means and variances of additional waiting time due to congestion. As potential demand increases, travel demand increases and causes the vehicles to operate at full capacity. Therefore, the means and variances of additional waiting time due to congestion increase, which directly influences the mean and variance of equilibrium effective travel cost. Since the equilibrium cost is higher, more routes are feasible and hence the size of the set of used routes increases.

By comparing cases 1 and 3 in Table 3, we can see that a higher value of congestion parameter, *n,* results in a higher mean and variance of additional waiting time due to congestion. More importantly, the set of used routes and the patronage of each line are highly affected by any change in the value of the congestion parameter *n*. In particular, when $n = 1$, there is no flow on route 2 and hence no patronage on transit lines L2, L3 and L4. Moreover, all passengers will take L1. However, when $n = 3$, the patronage of L1 is reduced by about half. The implication is that underestimating additional waiting time due to congestion can overestimate the patronage, which could have an adverse effect on the revenue and hence the profit.

By comparing cases 1 and 4 in Table 3, we can conclude that the equilibrium effective travel cost and the congestion level (reflected by the mean and variance of the additional waiting time due to congestion) are affected by the degree of risk aversion of passengers (reflected by the probability λ expressed as a percentage). More importantly, the degree of risk aversion of passengers has a major influence in determining the route flow pattern and hence the patronage of transit lines. In particular, we find that path 1 attracts about 7.5% less passengers when $\lambda = 99\%$ than when $\lambda = 50\%$ because path 1 has a higher travel cost variability than path 2 but this variability is only considered by the highly risk-aversive passenger with their $\lambda = 99\%$. This finding means that the patronage of line 1 would be

overestimated by 7.5 % if the risk aversion of passengers were not considered. This overestimation would have a profound impact on the revenue and hence the profit obtained from transit line L1. The private L1 service operator would lose money in the worst case if the operator set the fare by assuming that passengers ignored the variability of travel time while making their decisions.

To illustrate the effects of the value of the congestion parameter n and potential demand on the equilibrium effective travel cost, we varied n from 1 to 3, and for each value of n, we varied the potential demand from 400 to 2000 passengers/hour. Experimental runs were also carried out for three different cases of λ - 99%, 95% and 50%. By doing so, we can take into account the individual's degree of risk aversion. The results are plotted in Figures 4 and 5. From these figures, we can see that the equilibrium effective travel cost is monotonically increasing with potential demand under various values of n and λ (or ρ). As potential demand increases, the travel cost variability increases (as shown in Table 3), and hence travel cost margin and equilibrium effective travel cost increase. In addition, as λ increases, so do ρ, travel cost margin and equilibrium effective travel cost. However, in both figures, not all the curves are smooth, because the used path set is changing with potential demand. A kink can be observed at the boundary of potential demand, say 1200 passengers/hour for $n = 3$, where slightly increasing the potential demand increases the size of the used path set by at least one.

In order to study the effect of vehicle frequency on equilibrium effective travel cost, computational runs were carried out for three different scenarios. The first one is called the base scenario (denoted as b), which is studied with the same set of frequencies, the same potential demand and the same congestion parameter value as case 1 in table 3. All frequencies are reduced by 2 in the second scenario (denoted as b-2) and increased by 2 in the third scenario (denoted as b+2). The analysis was carried out for passengers with the three

risk aversion behaviors as before. The results are plotted in Figure 6, which shows that the changes in the frequencies under various risk aversion behaviors have a strong influence on the equilibrium effective travel cost. As each of the frequencies increases, there is a sharp reduction in the equilibrium effective travel cost and the reason behind this is that lower frequency results in higher mean and variance of waiting times.

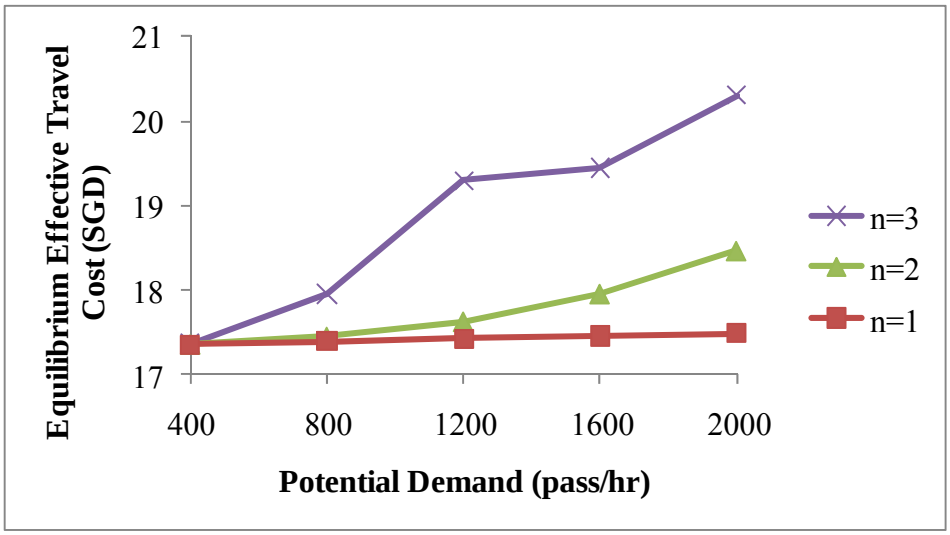


Figure 4: Equilibrium effective travel cost for various values of potential demand and congestion parameter

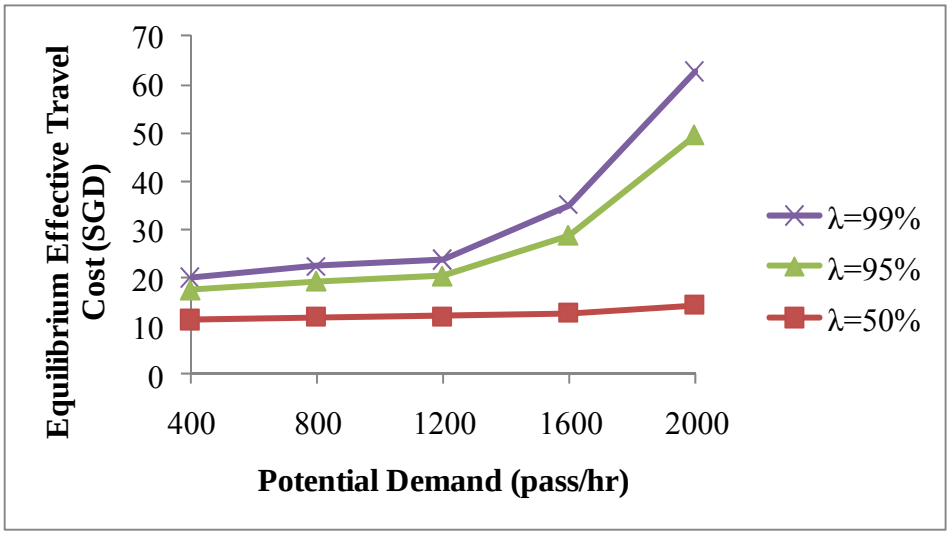


Figure 5: Equilibrium effective travel cost for various values of potential demand and degrees of risk aversion of passengers

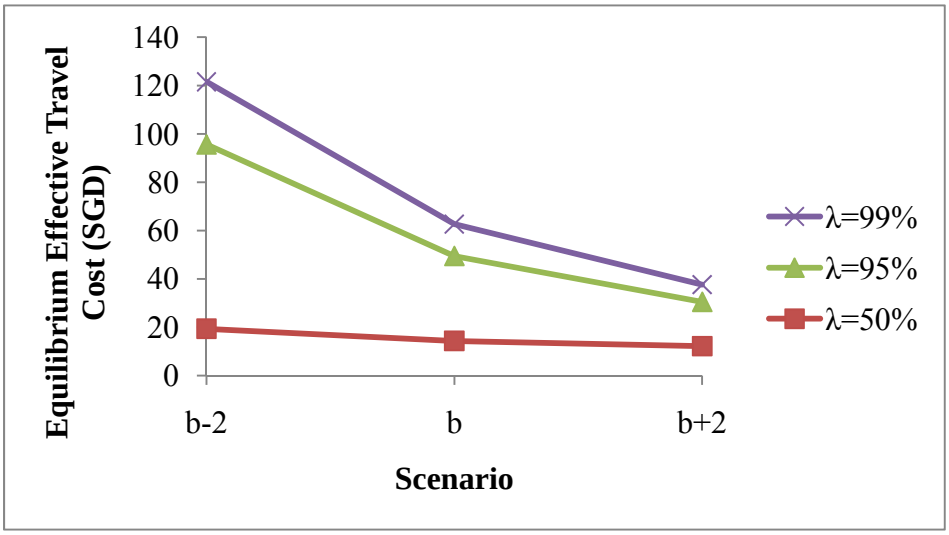
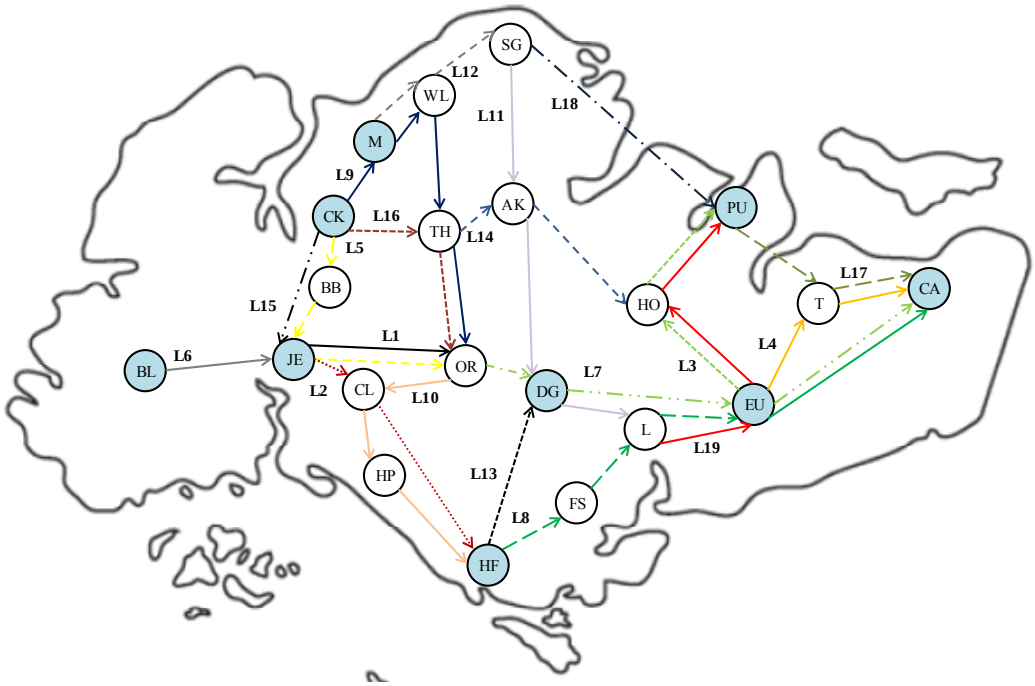


Figure 6: Influence of frequency on equilibrium effective travel cost under various degrees of risk aversion of passengers

Example 3 Effectiveness of the algorithm

To illustrate the effectiveness of the algorithm and its application to real networks, a larger network as shown in Figure 7 is used. The developed network is based on the Singapore bus network but only includes major stops and major services offered by Singapore Bus Services (SBS) Transit Limited. This network has 21 nodes, 19 lines and 19 OD pairs. The shaded nodes are origin or destination nodes. The corresponding network coded by route sections has 59 links. This network, we believe, captures all essential features of a large network including multiple OD pairs and many transit routes between each OD pair.

We consider the mean peak-hour frequency of each service as shown in the SBS Transit Limited webpage. The mean ρ value is estimated from the survey discussed in Example 1 and is found to be equal to 1.31. We do not have the actual demand data to calibrate the demand model. Therefore, we assume a potential demand pattern based on the given line capacity during the peak period, and carried out a sensitivity study on potential demand and the slope of the demand function. The values for remaining parameters follow those in the previous example.



Notations

AK	Ang Mo Kio	EU	Eunos	M	Marsiling
BB	Bukit Batok	FS	Fullerton Square	OR	Orchard
BL	Boon Lay	HF	Harbour Front	PU	Punggol
CA	Changi Airport	HO	Houngang	SG	Sembawang
CK	Choa Chu Kang	HP	Haw Par Villa	T	Tampines
CL	Clementi	JE	Jurong East	TH	Thomson
DG	Dhoby Ghaut	L	Lavender	WL	Woodlands

Figure 7: Simplified Singapore bus network

Table 4 demand sensitivity on solution speed

Demand factor df	1	2	3	4	5	6
Number of major iterations	17	22	16	171	215	477
CPU time (seconds)	4.66	6.23	4.44	52.81	66.56	105.88

Table 5: The effect of the slope of the linear demand function on solution speed

Slope χ_w	1	2	3	4	5	6
Number of major iterations	215	32	34	61	111	168
CPU time (seconds)	66.56	9.44	10.38	17.88	35.41	50.41

The proposed solution algorithm was coded in FORTRAN 90, and ran in a computer with an Intel Core Duo T2500 2.2GHz CPU, and a 1GB RAM. Table 4 shows the computation times required and the number of major iterations performed under different demand patterns, assuming $\chi_w = 1$. df is the demand factor which is used to scale up the potential demand of each OD pair in the base case. As you can see, a higher demand factor results in a longer computation time, meaning that a higher travel demand requires more computation time. This may be because a higher travel demand leads to more used paths and more interaction between different OD pairs, and hence reduces the speed of the convergence. Table 5 shows the effect of the slope of the demand function on the computation speed. In general, the slope greatly affects the convergence speed but we cannot conclude whether a large slope can reduce or improve the speed.

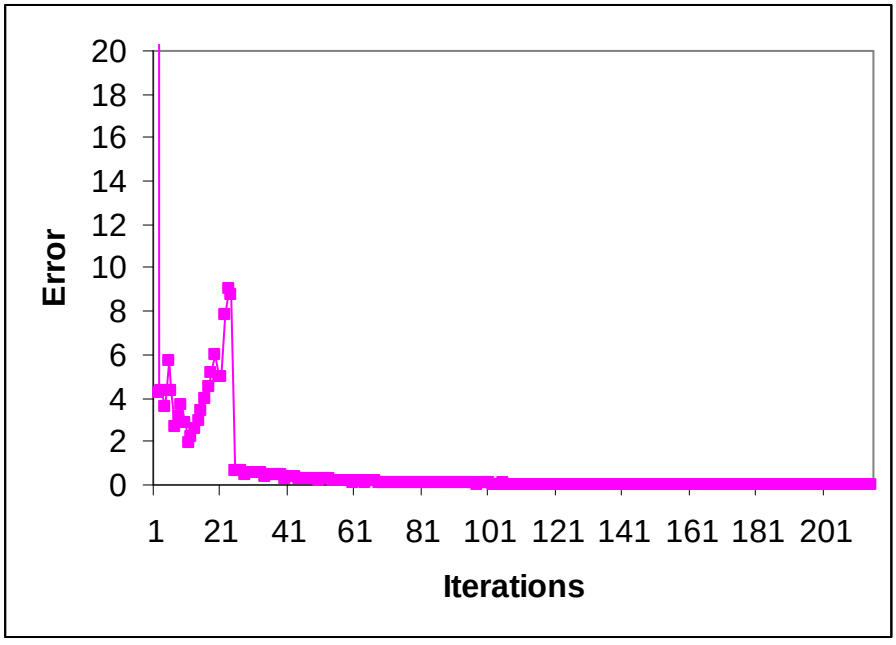


Figure 8: Errors over iteration

Figure 8 shows the convergence plot when $\chi_w = 1$ and $df = 5$. As revealed in this figure, the error G decreases over iteration on average but does not decrease monotonically probably because the mapping function in the NCP is not monotone with respect to path flow.

Indeed, the mathematical properties of this mapping function deserve a deep investigation because they have important implications on developing a convergent and efficient solution method for solving the proposed NCP formulation. We leave this to future studies.

5 CONCLUSIONS

In this paper, an NCP model is proposed for the reliability-based stochastic transit assignment problem. Compared with the frequency-based transit assignment model, the contribution of this paper is to propose a more realistic transit assignment model that

- considers both demand and supply uncertainties,
- captures risk-aversion behavior of passengers, and variabilities of in-vehicle travel time, waiting time, and congestion,
- can incorporate the proposed concept of travel cost budget, which is more general than the concept of travel time budget,
- has at least one solution,
- can separately model different effects of on-board passengers and passengers waiting at stops on congestion cost, and
- can be efficiently solved by the proposed column generation based solution method which can be applied to a realistic network.

Survey and numerical studies were also performed to validate the risk-aversion behavior of passengers, and to illustrate the properties of the problem and effectiveness of the proposed solution method. The results show that underestimating the congestion effect and ignoring the risk aversion behavior can overestimate the patronage of transit service. These findings have important implications on the profit of the operators involved and the development of transit network design models. The proposed model can be included in the transit network design model to determine the optimal service frequency and fare structure.

The proposed NCP model has been formulated for single-class passengers. Moreover, the perception errors on travel time and waiting time have not been considered yet. Based on the formulation proposed, it is not difficult to extend the formulation to consider the perception error, the distribution of the value of time and multiple user classes. This is left to future studies. Furthermore, 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. According to Nökel and Wekeck (2009), the dynamic passenger information systems give simultaneously the consecutive departure times for all lines serving a stop. There is no stochasticity involved in waiting time. The passengers can select the lines to minimize the sum of waiting time and in-vehicle travel time. The passengers do not need to board on the first arriving vehicle with sufficient capacity as assumed in this paper. This user behavior can be incorporated in our proposed framework in the future to correctly estimate the flow pattern and performance of the transit networks having such systems at some transit stops. A validation study can also be performed to test whether the extended approach can replicate the observed situation when the ridership data is available. In addition, the mathematical properties of the mapping function should be investigated in order to develop a convergent algorithm under a looser convergent requirement. Finally, based on the proposed framework, one can develop a transit network design model that captures risk-aversion behavior of passengers and develop a solution method based on heuristics such as tabu search (e.g., Fan and Machemehl, 2008), genetic algorithms (Ng *et al.*, 2009), and ant colony heuristics (e.g., Vitins and Axhausen, 2009) for the network design model.

REFERENCES

Abdel-Aty, M. A., Kitamura, R. & 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(1), 39-50.

Bell, M. G. H. & Cassir, C. (2002), Risk-averse user equilibrium traffic assignment: An application of game theory, *Transportation Research Part B*, 36(8), 671-681.

Cepeda, M., Corninetti, R., & 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(6), 437-459.

Chen, A., Lo, H.K., & Yang, H. (2001), A self-adaptive projection and contraction algorithm for the traffic assignment problem with path-specific costs. *European Journal of Operational Research*, 135(1), 27-41.

Chriqui, C. and Robillard, P. (1975) Common Bus Lines. *Transportation Science*, 9, 115-121.

Cominetti, R. & Correa, J. (2001), Common-lines and passenger assignment in congested transit networks, *Transportation Science*, 35(3), 250-267.

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

De Cea, J. & Fernández, E. (1993), Transit assignment for congested public transport systems: An equilibrium model, *Transportation Science*, 27(2), 133-147.

Fan, W. & Machemehl, R. B. (2008), Tabu search strategies for the public transportation network optimizations with variable transit demand, *Computer-Aided Civil and Infrastructure Engineering*, 23(7), 502-520.

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

Jackson, W. B. & Jucker, J. V. (1982), An empirical study of travel time variability and travel

choice behavior, *Transportation Science*, 16(4), 460-475.

Kulkarni, V.G. (1995), *Modeling and Analysis of Stochastic Systems*. Chapman & Hall. London.

Lam, W. H. K., Gao, Z. Y., Chan, K. S., & Yang, H. (1999), A stochastic user equilibrium assignment model for congested transit networks, *Transportation Research Part B*, 33(5), 351-368.

Lam, W. H. K., Zhou, J., & Sheng, Z. H. (2002), A capacity restraint transit assignment with elastic line frequency, *Transportation Research Part B*, 36(10), 919-938.

Larson, R. C. & Odoni, A. R. (1981), *Urban Operation Research* (available at http://web.mit.edu/urban_or_book/www/book/index.html).

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

Lei, Q. S. & Chen, J. (2004), An algorithm for transit assignment with elastic demand under capacity constraint, *Proceedings of the* 5^{th} *World Congress on Intelligent Control and Automation (WCICA)*, pp. 5245- 5247.

Lo, H. K., Luo, X. W., & Siu, B. W. Y. (2006), Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion, *Transportation Research Part B*, 40(9), 792-806.

Ministry of Manpower, Singapore (2009a), Ministry of Manpower | Earnings and Wages <http://www.mom.gov.sg/publish/momportal/en/communities/others/mrsd/statistics/Earnings_and_Wages.html> (accessed on 28 June 2009).

Ministry of Manpower, Singapore (2009b), Ministry of Manpower | Hour worked <http://www.mom.gov.sg/publish/momportal/en/communities/others/mrsd/statistics/Hours_Worked.html > (accessed on 28 June 2009).

Nagurney, A. (1999), Network Economics: A Variational Inequality Approach, Kluwer

Academic Publishers. Norwell, Massachusetts, USA.

Ng, M. W., Lin, D. Y., & Waller, S. T. (2009), Optimal long-term infrastructure maintenance planning accounting for traffic dynamics, *Computer-Aided Civil and Infrastructure Engineering*, 24 (7), 459-469.

Nguyen, S. & Pallottino, S. (1988), Equilibrium traffic assignment for large scale transit networks, *European Journal of Operational Research*, 37(2), 176-186.

Nökel, K. & Wekeck, S. (2009). Boarding and alighting in frequency-based transit assignment. Paper presented at 88th Annual Transportation Research Board Meeting, Washington D.C., January 2009.

Poon, M. H., Wong, S. C., & Tong, C. O. (2004), A dynamic schedule-based model for congested transit networks, *Transportation Research Part B*, 38(4), 343-368.

Schmöcker, J., Bell, M.G.H. & Kurauchi, F. (2008), A quasi-dynamic capacity constrained frequency-based transit assignment model, *Transportation Research Part B*, (42)10, 925-945.

Spiess, H. (1984) Contributions a La Theorie Et Aux Outils De Planification Des Reseaux De Transport Urbain. *Drpartement d'informatique et de recherché operationelle*. Universite de Montreal.

Spiess, H. & Florian, M. (1989), Optimal strategies: A new assignment model for transit networks, *Transportation Research Part B*, 23(2), 83-102.

Sumalee, A., Watling, D. P., & Nakayama, S. (2006), Reliable network design problem: the case with uncertain demand and total travel time reliability, *Transportation Research Record*, 1964, 81-90.

Sumalee, A., Tan, Z.J., Lam, W.H.K. (2009), Dynamic stochastic transit assignment with explicit seat allocation model, *Transportation Research Part B*, 43(8-9), 895-912.

Teklu, F. (2008), A stochastic process approach for frequency-based transit assignment with

strict capacity constraints, *Networks and Spatial Economics*, 8(2), 225-240.

Victoria Transport Policy Institute (2009), *Transportation Cost and Benefit Analysis II – Travel Time Costs*<www.vtpi.org/tca/tca0502.pdf > (access on 26 June 2009).

Vitins, B. J. & Axhausen, K. W. (2009), Optimization of large transport networks using the ant colony heuristic, *Computer-Aided Civil and Infrastructure Engineering*, 24(1), 1-14.

Yin, Y., Lam, W. H. K., & Ieda, H. (2004), New technology and the modelling of risk taking behaviour in congested road networks, *Transportation Research Part C*, 12 (3-4), 171-192.

APPENDIX: SURVEY QUESTION

Suppose you are a passenger going to A from B. Assume there are two bus lines (alternatives 1 and 2) connecting A and B. All the characteristics (fare, comfort, frequency, etc.) of the two bus lines are the same except the travel time (i.e., usual time) and possible delay in each scenario (a) to (f) shown in the table below. Please select one alternative for each scenario.

Paired Comparison for Determining a Respondent's Indifference Point

	Alternative 1	Alternative 2
(a) Usual time:	30 minutes	20 minutes
Possible delays:	None	5 minutes once a week
(b) Usual time:	30 minutes	20 minutes
Possible delays:	None	10 minutes once a week
(c) Usual time:	30 minutes	20 minutes
Possible delays:	None	15 minutes once a week
(d) Usual time:	30 minutes	20 minutes
Possible delays:	None	20 minutes once a week
(e) Usual time:	30 minutes	20 minutes
Possible delays:	None	25 minutes once a week
(f) Usual time:	30 minutes	20 minutes
Possible delays:	None	30 minutes once a week