Networks and Spatial Economics

Bounding the inefficiency of the reliability-based continuous network design problem under cost recovery

--Manuscript Draft--

Manuscript Number:	NETS-D-17-00106R2
Full Title:	Bounding the inefficiency of the reliability-based continuous network design problem under cost recovery
Article Type:	Manuscript
Keywords:	Inefficiency; price of anarchy; transport network design problem; reliability-based user equilibrium
Corresponding Author:	Wai Yuen Szeto The University of Hong Kong Hong Kong, HONG KONG
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	The University of Hong Kong
Corresponding Author's Secondary Institution:	
First Author:	Anny B. Wang
First Author Secondary Information:	
Order of Authors:	Anny B. Wang
	Wai Yuen Szeto
Order of Authors Secondary Information:	
Funding Information:	the University Research Committee of the University of Hong Kong (201711159034) — Dr. Wai Yuen Szeto the National Natural Science Foundation of China (71771194) — Dr. Wai Yuen Szeto the Guangzhou Science Technology and Innovation Commission (201707010292) — Dr. Anny B. Wang
Abstract:	This study defines the price of anarchy for general reliability-based transport network design problems, which is an indicator of inefficiency that reveals how much the design objective value exceeds its theoretical minimum value due to the risk averse and selfish routing behavior of travelers. This study examines a new problem, which is a reliability-based continuous network design problem under cost recovery. In this problem, the variations of system travel time and path travel times, the risk attitudes of the system manager and travelers, congestion toll charges, capacity expansions, and cost recovery constraint are explicitly considered. The design problem is formulated as a min-max problem with the reliability-based user equilibrium constraint. It is proved that the price of anarchy for this problem is bounded above, and the upper bound is independent of travel time functions, demands, and network topology. The upper bound is related to the travel time variations, the value of reliability, and the value of time.

Bounding the inefficiency of the reliability-based continuous network design problem under cost recovery

Anny B. Wang[1,2] and W.Y. Szeto[3&*]

[1]Guangzhou Municipal Engineering Design & Research Institute, Guangdong, P. R. China

[2]School of Civil Engineering and Transportation, South China University of Technology, Guangdong, P. R. China

[3]Department of Civil Engineering, The University of Hong Kong, Hong Kong, P. R. China

[&]The University of Hong Kong Shenzhen Institute of Research and Innovation, Shenzhen, P. R. China

[*] Phone: (852) 28578552, Fax: (852) 25595337, Email: ceszeto@hku.hk

Abstract

This study defines the price of anarchy for general reliability-based transport network design problems, which is an indicator of inefficiency that reveals how much the design objective value exceeds its theoretical minimum value due to the risk averse and selfish routing behavior of travelers. This study examines a new problem, which is a reliability-based continuous network design problem under cost recovery. In this problem, the variations of system travel time and path travel times, the risk attitudes of the system manager and travelers, congestion toll charges, capacity expansions, and cost recovery constraint are explicitly considered. The design problem is formulated as a min-max problem with the reliability-based user equilibrium constraint. It is proved that the price of anarchy for this problem is bounded above, and the upper bound is independent of travel time functions, demands, and network topology. The upper bound is related to the travel time variations, the value of reliability, and the value of time.

Keywords: Inefficiency, price of anarchy, transport network design problem, reliability-based user equilibrium

1 Introduction

The *price of anarchy* (PoA), which was first termed by Koutsoupias and Papadimitriou (1999), measures the inefficiency of the traffic assignment problem. It reveals how much the system performance measure would exceed its theoretical minimum value when travelers choose routes selfishly. The PoA for traffic assignment problems has received great research attention. Four major lines of research have arisen (Roughgarden and Tardos 2002; Chau and Sim 2003; Correa et al. 2004; Roughgarden 2005; Xiao et al. 2007; Han and Yang 2008; Han et al. 2008; Guo et al. 2010; Huang et al. 2011; Wang et al. 2014; Szeto and Wang 2015), which are based on four considerations: arc capacity constraints; demand and link travel time/cost functions; road pricing; and extensions of traditional user equilibrium principles and multiple user classes. The PoA for traffic assignment problems is well understood by scholars. However, the PoA for other problems, e.g., network design problems (NDPs), has rarely been studied.

The NDPs have broad definitions (Farahani et al. 2013). The most popular family of NDPs in the literature is the family of capacity expansion NDPs (Abdulaal and LeBlanc 1979; Dantzig et al. 1979; LeBlanc and Boyce 1986; Ben-Ayed et al. 1988; Friesz et al. 1993; Yang 1997; Yang and Bell 1998; Meng and Yang 2002; Chiou 2005; Szeto and Lo 2005; Szeto et al. 2010; Szeto et al. 2014), which optimizes the system performance measures of the road networks by determining the optimal capacity expansions (i.e., the additional capacities added to existing roads and/or the capacities of new roads) and the flow pattern (i.e., the traffic flow distribution in the road network). Some of these NDPs are also known as user equilibrium network design problems (UE-NDPs) because they capture the selfish routing behavior of travelers, which means that the flow pattern must satisfy the user equilibrium (UE) constraints. These NDPs also have one common feature—they assume that the travel demands and link capacities are deterministic.

In reality, there are uncertainties in the travel demands and road supplies due to day-to-day travel demand fluctuation, special events, bad weather, road accidents, road construction activities, etc. The demand and supply uncertainties lead to system travel time and path travel time variations, which cannot be ignored by the system manager and travelers. The reliability-based user equilibrium network design problems (RUE-NDPs) are developed based on the deterministic UE-NDPs by considering demand uncertainty and/or supply uncertainty. Chen et al. (2011) conducted a detailed review of the family of RUE-NDPs (Chootinan et al. 2005; Chen et al. 2007; Ng and Waller 2009; Sumalee et al. 2009; Yin et al. 2009; Chow and Regan 2011; Szeto and Wang 2016). Most existing studies focus on the modeling, solution methods, and applications of the capacity expansion RUE-NDPs. However, the PoA for the capacity expansion RUE-NDPs, which is an important indicator for evaluating how much the design objective function value exceeds its theoretical minimum value when travelers chose routes selfishly, has rarely been studied.

Szeto and Wang (2015) proposed the PoA for a capacity expansion RUE-NDP. Their study was the first attempt in the literature to examine the inefficiency of transport NDPs with capacity expansions. Szeto and Wang (2015) illustrated that the PoA for their proposed RUE-NDP reveals how much the system performance measure may exceed its corresponding theoretical minimum value due to the inefficient allocation of system resources (i.e., capacity expansions) and traffic flow, the latter of which is caused by the selfish routing behavior of travelers. They proved that the PoA has an upper bound, indicating that the inefficiency of the resource allocation of the network design is bounded above. The study of Szeto and Wang (2015) is far from complete. Firstly, they only considered one member of the capacity expansion RUE-NDP family. Their proposed PoA may not reflect the inefficiencies of resource allocations of the other RUE-NDPs that have different design objectives, decision variables, and constraints. Secondly, their study implicitly assumed that the RUE flow pattern is unique given the capacity expansions. Thirdly, most RUE-NDPs assume that the project cost does not exceed the available budget. However, the project cost can also be fully recovered by charging congestion tolls upon the travelers (Yang and Meng 2002; Lo and Szeto 2009). For RUE-NDPs that consider toll charges, the PoAs proposed by Szeto and Wang (2015) are not suitable. Thus, a general definition of the PoA for capacity expansion RUE-NDPs is required.

This study expresses the family of capacity expansion RUE-NDPs in a generalized model formulation and proposes a general definition of the PoA for the capacity expansion RUE-NDPs. This study then considers a specific problem, which is *a capacity expansion RUE-NDP* under cost-recovery that considers supply uncertainty and road tolls. The problem is formulated as a min-max problem. The *min*-level problem aims to minimize the largest total system travel cost budget (TSTCB) plus the project cost. The TSTCB is a variant of the total system travel time budget and consists of the monetary cost of mean total system travel time and an extra cost associated with system travel time reliability. The *max*-level problem aims to determine the worst-case flow pattern that gives the largest TSTCB plus the project cost. The self-routing behavior and risk attitudes of travelers are captured by the reliability-based user equilibrium (RUE) constraints. In addition, travelers are charged with congestion tolls, which are used to recover the project cost. To guarantee that the project is self-financing or even profitable, a cost recovery constraint is incorporated. Based on the proposed model, this study proposes a novel approach to derive the analytical formula for an upper bound of the PoA.

The contributions of this study are as follows:

- We propose a general definition of the PoA for capacity expansion RUE-NDPs to measure the inefficiency of the reliability-based transport NDPs with capacity expansion and cost recovery;
- We propose a new NDP, namely capacity expansion RUE-NDP under cost recovery, in which the project cost is fully recovered by charging travelers with congestion tolls. It is formulated by a min-max approach; and
- It derives an analytical bound of the PoA of the proposed capacity expansion RUE-NDP under cost recovery.

The key findings regarding the upper bound of the PoA for the proposed RUE-NDP include the following:

- The upper bound depends on the travel time variations, the value of travel time, the value of reliability for system travel time, and the value of reliability for path travel time;
- The upper bound is independent of travel time functions, demands, and network topology; and
- The upper bound equals one if there are no travel time variations or/and the system manager and travelers are both risk-neutral, indicating that the PoA also equals one.

This paper is organized as follows. In Section 2, we express the family of capacity expansion RUE-NDPs in a generalized model formulation and propose a general definition of the PoA for the capacity expansion RUE-NDPs. In Section 3, we describe our new problem. In Section 4, we examine the PoA for the studied problem and evaluate its upper bound. In Section 5, we provide a concluding remark and discuss the future research directions.

2 PoA for the capacity expansion RUE-NDPs

Consider a road network with topology $G(N, A)$, in which N is a finite set of nodes and A is a finite set of directed links. The nodes represent existing or candidate intersections. The directed links represent roads whose existing capacities are to be expanded or whose capacities are to be determined. The network has multiple origin-destination (O-D) pairs that define where the travelers are from and where they head to. Each O-D pair is associated with its travel demand, which is the number of travelers between the origin and the destination per hour.

For the clarity of the presentation, the main notations are defined and introduced in Table 1.

Table 1. Notations

$\mathbb{R}$	The set of real numbers
$\mathbb{R}_+$	The set of positive real numbers
RS	The set of O-D pairs in the road network
P	The set of all possible paths connecting different O-D pairs in the road network; its size is denoted by $m \in \mathbb{R}_+$
P_{rs}	The set of all possible paths connecting O-D pair rs, $rs \in RS$
d_{rs}	The positive travel demand or mean travel demand between O-D pair $rs \in RS$
$\mathbf{d}$	The vector of travel demands/mean travel demands between all O-D pairs $\left(d_{rs}\right)_{rs\in RS}$
δ_p^a	The link-path incidence indicator, which equals one if link $a \in A$ is on path $p \in P$, and equals zero otherwise
f_p	The non-negative flow or mean flow on path $p \in P$
$\mathbf{f}$	The vector of path flows or mean path flows $\left(f_p\right)_{p\in P}$
Ω_f	The set of feasible path flow patterns that satisfy the path-flow demand conservation constraints and non-negativity constraints: $\Omega_f = \left\{ \mathbf{f} \,\middle|\, \sum_{p\in P_{rs}} f_p = d_{rs}, \forall rs \in RS;\ f_p \geq 0, \forall p \in P \right\}$
v_a	The non-negative flow or mean flow on link $a \in A$
$\mathbf{v}(\mathbf{f})$	The vector of link flows or mean link flows in the road network $\left(v_a\right)_{a\in A}$ with $v_a = \sum_{p\in P} f_p \delta_p^a$, $\forall a \in A$, $\mathbf{f} = \left(f_p\right)_{p\in P} \in \Omega_f$
y_a	The design variable, which is the capacity of a new link $a \in A$ or the link capacity expansion of an existing link $a \in A$
u_a	The upper bound of y_a, $a \in A$
Ω_y	The set of feasible link capacities or link capacity expansions: $\Omega_y = \left\{ \mathbf{y} \middle| 0 \leq y_a \leq u_a, \forall a \in A \right\}$

$\mathbf{y}$	The vector of the capacities of new links or link capacity expansions of existing links $\left(y_a\right)_{a\in A}$
$t_a(v_a, y_a)$	The mean link travel time function of link $a \in A$ in terms of its link flow and link capacity (expansion)
$\mathbf{t}$	The vector of mean link travel time functions $\left(t_a\right)_{a\in A}$
$\boldsymbol{\sigma}$	The covariance matrix, which contains all the link travel time variances and link travel time covariances

2.1 Generalized model formulation of capacity expansion RUE-NDPs

The capacity expansion RUE-NDPs have various input information known as the *design instances*. A design instance is described by the general form $\left(G, \mathbf{d}, \mathbf{t}, \theta\right)$, in which $\mathbf{d}$ and $\mathbf{t}$ are defined in Table 1, and θ stands for any additional and essential information related to the RUE-NDP. θ can be a scalar, a vector, or a set of vectors. For example, θ may include the project budget and the travel time variation related information.

A capacity expansion RUE-NDP is formulated as a bi-level mathematical optimization problem with decision variables, constraints, and an objective function.

The decision variables include the vector of capacity expansions (i.e., $\mathbf{y}$). The capacity expansions include the additional capacities added to existing roads and/or the capacities of new roads. Other decision variables include the path flow pattern $\mathbf{f}$. Note that the link flow pattern $\mathbf{v}(\mathbf{f})$ is dependent on the path flow pattern $\mathbf{f}$. Thus, the link flows are dependent variables. In an RUE-NDP, $\mathbf{y}$ and/or $\mathbf{f}$ may be random variables. The decision variables in the RUE-NDP are commonly the *mean* capacity expansions and *mean* link flows. In addition, in some NDPs (e.g., Szeto and Lo 2005, Lo and Szeto 2009), the travelers are charged with road tolls. The link tolls are commonly dependent variables whose values depend on the link flows. For convenience, we denote any auxiliary decision variables as a vector $\mathbf{w}$ whose feasible set is described by a non-empty set X_0.

The constraints of a capacity expansion RUE-NDP include the *feasibility constraints*, i.e., the path flow-demand conservation constraints, the link-path flow conservation constraints, the non-negativity constraints of path flows and capacity expansions, and the feasibility constraints of the auxiliary decision variables. These constraints are implicitly captured by the non-empty sets Ω_f, Ω_y, X_0, and the definition of $\mathbf{v}(\mathbf{f})$. Specifically, the constraint set Ω_y restricts which links can have capacity changed and which new links can be added, and hence any strategy would be embodied in this constraint set and in other additional constraints. Most importantly, the RUE-NDP incorporates a set of non-linear inequalities and equalities known as the *RUE constraints*. The constraints capture the self-routing behavior of travelers or the risk attitudes of travelers. Apart from the feasibility constraints and the RUE constraints, the RUE-NDP might also have other related constraints, such as the *budget constraint*, which guarantees that the project cost is not larger than the project budget. If travelers are charged with road tolls, the budget constraint can be replaced by the *cost recovery constraint* (Lo and Szeto 2009), which guarantees that the project cost is not larger than the total toll revenue collected from the travelers.

The *system performance measures* of the road network include the consumer surplus (Yang 1997), the reserve capacity (Yang and Bell 1998), the total vehicle miles (Friesz et al. 1993), the sum of total system travel time and construction cost (Chiou 2005), and the total system travel time/cost (Meng and Yang 2002). The objective function of an RUE-NDP includes the mean system performance measure (Chow and Regan 2011), the sum of the mean and (weighted) variance/standard deviation of the system performance measure (Ng and Waller 2009; Sumalee et al. 2009; Szeto and Wang 2016), and the worst-case value of the system performance measure (Yin et al. 2009). The objective function of the RUE-NDP is commonly a continuous function in terms of the decision variables, denoted as $Z(\cdot)$. In this study, we assume that the objective function value is dependent on the link flow pattern (or path flow pattern) and the capacity expansions, and is independent of the auxiliary decision variables.

For most existing capacity expansion RUE-NDPs, the objective is to minimize the objective function. However, such a design objective is optimistic when there are multiple link flow patterns for a given $\mathbf{y}$ (e.g., Liu et al., 2017). In fact, Wang and Szeto (2018) proved that the RUE link flow pattern is unique when two conditions hold: 1) the path travel costs are monotone in terms of path flows; 2) the link travel cost is a bijective function of link flow. If the RUE link flow pattern is non-unique, the actual RUE flow pattern after the implementation of the capacity expansions may be different from the design RUE flow pattern, yielding a worse system performance than what the system manager expected. To deal with this practical issue, we consider that the system manager (or a risk-averse system manager) aims to minimize the *worst* possible value of the objective function over $\mathbf{y}$, i.e., minimizing the maximum value of the objective function over $\mathbf{y}$.

Based on the above, we express the capacity expansion RUE-NDPs as the following *general* non-linear constrained optimization problem:

$$\min_{\mathbf{y}} \max_{\mathbf{f}} \; Z(\mathbf{v}(\mathbf{f}),\mathbf{y}), \tag{1}$$

subject to the *RUE constraints*:

$$\bar{g}_i(\mathbf{v}(\mathbf{f}),\mathbf{y},\mathbf{w}) \le 0, \; i = 1,2...,m \in \mathbb{R}^+, \tag{2}$$

$$g_i(\mathbf{v}(\mathbf{f}),\mathbf{y},\mathbf{w}) = 0, \; i = 1,2...,m \in \mathbb{R}^+; \tag{3}$$

the *feasibility constraints*:

$$\mathbf{f} \in \Omega_f, \mathbf{y} \in \Omega_y, \tag{4}$$

$$\mathbf{w} \in X_0; \tag{5}$$

and *other relevant sets of constraints (e.g., budget constraints or cost recovery constraints)*:

$$h_i(\mathbf{v}(\mathbf{f}),\mathbf{y},\mathbf{w}) \le 0, \; i = 1,2...,n \in \mathbb{R}_+, \tag{6}$$

where $\bar{g}_i$, g_i, and h_i are all functions of $\mathbf{v}(\mathbf{f})$, $\mathbf{y}$, and $\mathbf{w}$. In constraints (2) and (6), m is the total number of paths and n is the total number of additional constraints.

If the objective function $Z(\mathbf{v}(\mathbf{f}),\mathbf{y})$ is continuous, the set described by constraints (4), (5), and (6) is non-empty, and an RUE link flow pattern exists and satisfies the equilibrium constraints (2)-(3), then the optimization problem (1)-(6) has at least one optimal solution, denoted as $(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}},\tilde{\mathbf{w}})$. For any $\mathbf{y}$, if the RUE link flow pattern is unique, the problem (1)-(6) is equivalent to $\min_{\mathbf{f},\mathbf{y}} Z(\mathbf{v}(\mathbf{f}),\mathbf{y})$ subject to (2)-(6).

Remark. The generalized model formulation (1)-(6) can also be used to express the UE-NDPs, because the UE-NDPs are special cases of RUE-NDPs in which the travel time variations are zero and/or the travelers and the system manager are both risk-neutral.

2.2 General definition of the PoA for capacity expansion RUE-NDPs

Firstly, to show the rationality of defining the PoA for capacity expansion RUE-NDPs, we quote the statement of Roughgarden (2005): "The price of anarchy can be defined much more generally; indeed, the concept makes sense for every application possessing an objective function and a notion of equilibrium".

Secondly, we identify the theoretical minimum objective function value when all the travelers willingly choose paths to minimize the objective function value. The minimum objective function value is obtained by minimizing $Z(\mathbf{v}(\mathbf{f}),\mathbf{y})$ subject to the feasibility constraints (4), (5) and (6). The problem is referred to as a capacity expansion Reliability-based System Optimum NDP (RSO-NDP) and it is expressed as the following general non-linear minimization problem:

$$\min_{\mathbf{f}\in\Omega_f,\mathbf{y}\in\Omega_y} Z(\mathbf{v}(\mathbf{f}),\mathbf{y}). \tag{7}$$

The solution which yields the minimum objective function value is called the *system optimal solution*, and we denote it as $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*,\mathbf{w}^*)$. To differentiate the system optimal solution and the optimal solution to the RUE-NDP (i.e., $(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}},\tilde{\mathbf{w}})$), we call the latter the *equilibrium solution*.

Conceptually, the PoA is the worst-possible ratio between the objective function value of an equilibrium solution and that of a system optimal solution. A formal mathematical definition is given as follows.

i. Given an instance $(G,\mathbf{d},\mathbf{t},\theta)$ admitting a system optimal solution $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*,\mathbf{w}^*)$ and an equilibrium solution $(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}},\tilde{\mathbf{w}})$, the *PoA of* $(G,\mathbf{d},\mathbf{t},\theta)$ is

$$\rho(G,\mathbf{d},\mathbf{t},\theta) = Z(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}}) \big/ Z(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*). \tag{8}$$

ii. Denote the set of design instances that have some common features as I, e.g., the set of instances whose travel time functions are all Bureau of Public Road type link performance functions. The *PoA of* I is

$$\rho(I) = \sup_{(G,\mathbf{d},\mathbf{t},\theta)\in I} \rho(G,\mathbf{d},\mathbf{t},\theta). \tag{9}$$

Remark. The mathematical definition of the PoA may take different forms. For example, the pioneer study (Roughgarden 2005) included the two terms $(G,\mathbf{d},\mathbf{t},\theta)$ and I in the definition of the PoA for the classical traffic assignment problem (see Definition 2.3.1 (a) and (b) in his study), whereas some studies omitted them. In this study, we take the study of Roughgarden (2005) as the reference and include the two terms in the definition of the PoA for capacity expansion RUE-NDPs.

The PoA reflects the inefficiency of equilibrium solutions to the RUE-NDPs. The inefficiency refers to two aspects, which are both caused by the selfish-routing behavior of travelers: 1) the traffic flow distribution is not the best; and 2) the allocation of resources (capacity expansion) is not the best. In practice, the PoA is an economic evaluation index, based on which the system manager can quickly determine the relative reduction of system performance induced by the selfish-routing behavior of travelers brings to the transport network design. The PoA is a ratio and it is intuitively larger than one. A smaller PoA value indicates that the efficiency loss is less, and vice versa.

The $\rho(G,\mathbf{d},\mathbf{t},\theta)$ in (8) reflects the *exact inefficiency* of an equilibrium solution to the RUE-NDP with instance $(G,\mathbf{d},\mathbf{t},\theta)$. The $\rho(I)$ in (9), on the other hand, reveals the worst-case inefficiency of equilibrium solutions to the RUE-NDP with instances that share some common feature.

The PoA for the capacity expansion RUE-NDPs proposed in this study differs from the PoAs proposed by Szeto and Wang (2015). The PoAs proposed by Szeto and Wang (2015) are defined for the RUE-NDP that must satisfy the following conditions: 1) the lower level reliability-based user equilibrium flow patterns must be unique; 2) the decision variables are merely link capacity additions; 3) the design objective functions are total system travel time and total system travel time budget; 4) the RUE-NDP only considers supply uncertainty; and 5) the reliability-based user equilibrium problem adopts the travel time budget approach (Shao et al. 2006). The PoA proposed in our study, on the other hand, is defined for RUE-NDPs that satisfy less restrictive conditions. Firstly, the RUE-NDPs may have additional decision variables such as the road tolls. It allows the system manager to evaluate the impacts of the additional decision variables on the inefficiency of resource allocation. Secondly, apart from the classic system performance measure, which is the cost of system travel time, the objective functions may also include the cost of travel time reliability, environmental cost, construction cost, etc. It allows the system manager to evaluate the inefficiency of resource allocation with respect to different additional considerations such as travel time uncertainty, environmental impacts, and project cost, etc. Thirdly, the RUE-NDPs may incorporate additional constraints (e.g., the cost recovery constraint), which allows the system manager to evaluate the inefficiency of resource allocation when there are additional constraints to consider. Fourthly, the RUE-NDP may consider demand uncertainty/supply uncertainty or both, allowing the system manager to evaluate the inefficiency of resource allocation when the demand and/or supply are random variables. Finally, the lower level RUE problem of the RUE-NDPs may be formulated by other approaches. It allows the system manager to consider different types of RUE problems such as the mean-excess travel time (Chen and Zhou 2010) RUE problem, the stochastic dominance RUE problem (Wu and Nie 2011), and the non-expected route choice problem (Ji et al. 2017), etc.

To further illustrate the PoA in detail, we consider a specific problem proposed in the following, which is a capacity expansion RUE-NDP under cost recovery. The problem determines the capacities of the new roads in a road network under supply uncertainty and is formulated as a min-max problem. The travelers are charged with congestion tolls after the road network is built and put into usage. The construction cost of the road network is fully recovered from toll charges.

3 Reliability-based capacity expansion NDP under cost recovery: Min-max formulation

3.1 Objective function

Consider that the system manager designs which roads are expanded and/or built. Moreover, the manager considers the effect of supply uncertainty in the network design: the *actual link capacities may degrade from their design values (Szeto and Wang, 2015, 2016; Zhao et al., 2018) and the actual link free flow travel times may deviate from their pre-assumed values derived from maximum allowed speeds (Szeto and Wang, 2015, 2016).* The demands and the link flows are deterministic. The travel time on a link $a \in A$ (denoted by T_a) is thus modeled as a random variable.

From the system manager's perspective, his/her primary design objective is to minimize the total system travel time (TSTT). The TSTT equals the sum of the travel times experienced by all travelers. Thus, the TSTT is a compound random variable. We denote it as $TSTT$, and it equals $TSTT = \sum_{a \in A} T_a v_a$.

The expectation and standard deviation of the compound random variable $TSTT$ can be obtained by the following operations:

$$E\left[TSTT\right] = E\left[\sum_{a \in A} T_a v_a\right] = \sum_{a \in A} E\left[T_a\right] v_a ,$$

$$\sigma\left[TSTT\right] = \sigma\left[\sum_{a \in A} T_a v_a\right] = \left(\sum_{a \in A} \sigma^2\left[T_a\right] v_a^2 + \sum_{a \in A} \sum_{a' \in A, a \neq a'} v_a v_{a'} Cov\left[T_a, T_{a'}\right]\right)^{1/2} .$$

Commonly, the mean link travel time $E\left[T_a\right]$ of link $a \in A$ is predicted by its link travel time function $t_a(v_a, y_a)$. We assume that $t_a(v_a, y_a)$ is a *bijective* function with respect to its link flow given the link (additional) capacity. The link travel time function is monotone increasing and differentiable with respect to v_a, and monotone decreasing and differentiable with respect to y_a. We also assume that the link travel time variance $\sigma^2\left[T_a\right]$ and the travel time covariances $Cov\left[T_a, T_{a'}\right]$, $a' \in A$, $a' \neq a$ are finite. The explicit functional forms of the travel time variances depend on the link travel time functions and the assumed distributions of link free flow travel times and random link capacities.

Szeto and Wang (2015, 2016) proposed the concept of *total system travel time budget*, which simultaneously captures the mean and variation of TSTT, and is defined as:

Total system travel time budget = mean total system travel time + safety margin .

However, the system performance measure with a time unit is less preferable in practice because the investment parties are more concerned with the project cost rather than the TSTT itself. The system manager should consider the concerns of these parties. However, the TSTT cannot be directly combined with the project cost. Similarly, the total system travel time budget is also not a suitable indicator because it cannot be directly combined with the project cost. Thus, a similar concept to the total system travel time budget— *the TSTCB*—is proposed:

Total system travel cost budget = monetary value of mean total system travel time + monetary value of system travel time reliability.

The monetary value of mean TSTT can be obtained by multiplying the mean TSTT by a positive coefficient representing the value of time (VOT) for mean travel time:

$$\text{monetary value of mean TSTT} = \text{VOT} \cdot \text{mean TSTT},$$

in which the VOT is obtained by calibration using the survey data. The VOTs of road networks in different areas (e.g., cities, country regions, or countries) are different. Relevant studies on the VOT include the studies of Small and Yan (2001), Brownstone and Small (2003), and Tilahun and Levinson (2009).

The VOR converts a measure of travel time reliability into the monetary value of travel time reliability. The monetary value of travel time reliability can be obtained by

$$\text{monetary value of travel time reliability} = \text{VOR} \cdot \left(\text{measure of travel time reliability}\right).$$

The measures of travel time reliability include the difference between the 90th and 50th percentile travel time, the standard deviation of travel time, the difference between the actual late arrival and the usual travel time, and the difference between the early/late arrival time and the preferred arrival time. Given different measures of travel time reliability, the VORs are different. In this study, the standard deviation of TSTT is adopted as the measure of travel time reliability and used in the TSTCB.

Mathematically, the *TSTCB* is defined as follows:

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}} = R^{\mathrm{t}} \sum_{a\in A} E\left[T_a\right] v_a + R^{\mathrm{s}} \sqrt{\sum_{a\in A} \sigma^2\left[T_a\right] v_a^2 + \sum_{a\in A} \sum_{a'\in A, a\neq a'} v_a v_{a'} Cov\left[T_a, T_{a'}\right]},$$

in which R^{t} is the VOT for mean TSTT and R^{s} is the VOR for total system travel time.

There are no references for R^{s}. The report by Concas and Kolpakov (2009) only summarized the VORs for path travel time obtained by different studies. Nevertheless, the statistical methods used to calibrate the VOR for path travel time in that studies can also be used to calibrate R^{s}. Similar to the fact that the VOR for path travel time is dependent on the risk aversion of the travelers, R^{s} is related to the risk-aversion of the system manager. A larger R^{s} indicates that the system manager is more risk averse, and vice versa. The R^{s} equals zero if the system manager is risk neutral or/and considers that there is no monetary value in the reliability of TSTT.

As discussed before, apart from optimizing the system performance measure, the project cost is also an important consideration for the system manager. To formulate it, the annual cost of a link $a \in A$, denoted as $I_a(y_a)$, is introduced:

$$I_a(y_a) = \kappa_a y_a, \ \kappa_a > 0, \ \forall a \in A,$$

where the constant κ_a represents the annual cost per unit of (additional) capacity of link a. The annual cost per unit of (additional) capacity of a link $a \in A$ (i.e., κ_a) captures two factors: the annualized construction cost per unit of (additional) capacity and the annual maintenance cost per unit of (additional) capacity. The definition of $I_a(y_a)$ is based on two assumptions: 1) There is a constant return to scale in road construction, and 2) the maintenance/operation cost per unit of (additional) capacity is constant. The project cost equals the annual overall costs associated with the construction and maintenance of the road network, and we call it the *investment cost* (IC), which is

$$IC(\mathbf{y}) = \sum_{a\in A} I_a(y_a) .$$

From the system manager's perspective, the design objective is to minimize the sum of the TSTCB and IC, i.e.,

$$\min_{\mathbf{y}\in\Omega_y,\mathbf{f}\in\Omega_f} TSTCB_{R^t,R^s}(\mathbf{v}(\mathbf{f}),\mathbf{y}) + IC(\mathbf{y}) . \tag{10}$$

Note that if IC is not considered, then the above optimization model belongs to the family of mean-standard deviation models (e.g., Lo et al., 2006; Khani and Boyles, 2015; Wu, 2015).

3.2 RUE constraints with link marginal mean cost tolls

The travelers' selfish-routing and risk-adverse behaviors are captured by the RUE constraints. The RUE constraints are developed from Wardrop's first principle (Wardrop 1952), which states that a traveler always chooses a path that minimizes his/her own travel time. The travel time of a path equals the sum of the link travel times of all links on that path. Because the link travel times are all random variables, the path travel time, denoted as Q_p, $p\in P$, is also a random variable and expressed as

$$Q_p = \sum_{a\in A} T_a \delta_p^a,\ \forall p\in P .$$

The mean path travel time $E\left[Q_p\right]$, denoted as q_p, is $q_p = \sum_{a\in A} t_a \delta_p^a,\ \forall p\in P$.

When faced with travel time uncertainties, travelers often depart early and reserve extra time for their trips to avoid late arrivals. The risk-averse behavior of travelers is well known and many approaches extended from Wardrop's principle have been proposed to capture it. Among them, the path travel time budget (TTB) approach (Lo et al. 2006) is frequently adopted. The TTB approach assumes that each traveler selects a path with the minimum path TTB. The TTB is commonly defined as the sum of the mean path travel time and the weighted path travel time standard deviation.

Similar to the total system travel time budget, the path TTB also has a time unit. A variant of the TTB is the path travel cost budget, which has a cost unit and is defined as follows.

Path travel cost budget = monetary value of mean path travel time +
monetary value of path travel time reliability.

Similar to the TSTCB, the monetary values of mean path travel time and path travel time reliability can be obtained by the following operations:

monetary value of mean path travel time $= \mathrm{VOT}\cdot q_p$, and

$$\text{monetary value of path travel time reliability} = \text{VOR} \cdot \left(\text{measure of path travel time reliability}\right),$$

in which the measure of path travel time reliability is the path travel time standard deviation. Based on the above, the *path travel cost budget* b_p, $\forall p \in P$ is

$$b_p = R^{\mathrm{t}} \cdot q_p + R^{\mathrm{u}} \cdot \sigma\left[Q_p\right],$$

in which $R^{\mathrm{t}} > 0$ is the VOT for mean path travel time and $R^{\mathrm{u}} \geq 0$ is the VOR for *path travel time.*

The VOT for mean path travel time and the VOT for mean total system travel time are consistent with each other, which are both R^{t}. As the measure of path travel time reliability is the path travel time standard deviation, the values for R^{u} can be found in the study of Concas and Kolpakov (2009).

It is assumed that all travelers are charged with congestion tolls because congestion toll charging has been adopting to mitigate congestion and improve system performance in reality. For a road network without uncertainties, *link marginal cost tolling* is one of the well-known tolling strategies for driving a UE flow pattern towards a flow pattern that yields a better system performance (Yang and Meng, 2002), and it is defined as *the product of the link flow and the first-order derivative of the link travel time function with respect to the link flow, assuming that the value of time is one*. For a road under supply uncertainty, however, because of the travel time variations, it is unclear whether charging the corresponding link marginal cost tolls will lead to an improvement in *TSTCB*. It only improves the *mean* TSTT. Nevertheless, this study assumes that the system manager adopts the link marginal cost tolls called *link marginal mean cost tolls* in a road network under supply uncertainty. The *link marginal mean cost toll* on link *a* is denoted by τ_a and defined by

$$\tau_a = R^{\mathrm{t}} v_a \cdot dt_a(v_a, y_a) / dv_a, \ \forall a \in A.$$

For a traveler, the *generalized path travel cost budget*, denoted by $\tilde{b}_p$, $\forall p \in P$, is

$$\tilde{b}_p = b_p + \sum_{a \in A} \delta_p^a \tau_a(v_a, y_a).$$

It is assumed that the travelers acquire the expectations and variabilities of path travel times, the VOT for path travel time, the VOR for path travel time standard deviation, and the link marginal mean cost tolls based on their experiences and factor this piece of information into their route choice considerations in the form of a generalized path travel cost budget. All travelers select routes to minimize their generalized path travel cost budgets. The long-term equilibrium is reached only if the generalized path travel cost budgets of all used routes are not higher than those of unused routes. The RUE flow path pattern $\mathbf{f}^{RUE} = (f_p^{RUE})_{p \in P}$ and the corresponding link flow pattern $\mathbf{v}^{RUE}(\mathbf{f}^{RUE})$ must satisfy the following *RUE constraints*:

$$f_p^{RUE}\left(\tilde{b}_p(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}), \mathbf{y}^{RUE}) - w_{rs}\right) = 0, \ \forall p \in P_{rs}, \ \forall rs \in RS, \tag{11}$$

$$\tilde{b}_p(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}), \mathbf{y}^{RUE}) - w_{rs} \geq 0, \ \forall p \in P_{rs}, \ \forall rs \in RS, \tag{12}$$

where w_{rs} is the minimum generalized path travel cost budget for O-D pair $rs \in RS$, and $\mathbf{y}^{RUE}$ is an optimal capacity solution to be determined. Denote $\mathbf{w} = \left(w_{rs}\right)_{rs \in RS}$ and it is the vector of auxiliary decision variables that must be non-negative, i.e.,

$$\mathbf{w} \geq \mathbf{0}. \tag{13}$$

Denote the standard deviation of the path travel cost as ς_p, $p \in P$. Unlike the mean link travel times, the mathematical property of ς_p is not known until the explicit formulation of link travel time standard deviations and travel time covariances are known. Without the loss of generality, we assume that the mapping $\boldsymbol{\varsigma} = \left(\varsigma_p\right)_{p \in P}$ is monotone with respect to the path flow pattern $\mathbf{f}$. Then, the path travel cost budgets are monotone with respect to the path flows. In addition, the mean link travel times are bijective functions of link flows. Following the proofs of Wang and Szeto (2018), the minimum path travel cost budgets, the monetary values of mean link travel times, and the RUE link flow pattern at equilibrium are unique. The RUE path flow pattern, on the other hand, is non-unique.

The generalized path travel cost budget includes the link marginal mean cost tolls. One of the purposes of charging link marginal mean cost tolls upon the travelers is to recover the IC. To check whether the total toll revenue collected from travelers covers the IC or not, the concept of the degree of cost recovery is introduced and defined in the next section.

3.3 Cost recovery constraint

A notion, namely the *degree of cost recovery***,** denoted by η_τ, is defined as

$$\eta_\tau = \left(\boldsymbol{\tau}^{\mathrm{T}} \cdot \mathbf{v}(\mathbf{f})\right) \Big/ \left(\boldsymbol{\kappa}^{\mathrm{T}} \cdot \mathbf{y}\right),$$

where $\boldsymbol{\kappa}$ is the vector of the annual costs per unit of (additional) capacity defined in Sub-section 3.1.

The ratio defined in the above has been mentioned and adopted by Szeto and Lo (2008). The *degree of cost recovery* is an important indicator showing how profitable a toll scheme is. The project is profitable if η_τ is larger than one. The project is cost-recovery if η_τ is larger than or equal to one. The project is self-financing if η_τ exactly equals one. If η_τ is smaller than one, the total revenue collected from travelers cannot cover the IC, which means that the toll scheme $\boldsymbol{\tau}$ is not satisfactory from an investment perspective.

To guarantee that at an optimal design, the IC is fully covered by the total toll revenue collected from travelers, a cost recovery constraint is incorporated into the design problem. That is, the degree of cost recovery must be larger than or equal to one:

$$\eta_\tau \geq 1. \tag{14}$$

3.4 Model formulation

One possible way to depict the capacity expansion RUE-NDP under cost recovery is that it minimizes the objective function (i.e., (10)) subject to the RUE constraints (i.e., (11) and (12)), the cost recovery constraint (i.e., (14)), and the feasibility constraints of the decision variables. However, for a given $\mathbf{y}$, there might be multiple RUE link flow patterns that satisfy the RUE constraints. The objective (10) naturally selects the solution that has the minimum objective function value. In practice, the actual RUE flow pattern may deviate from the design (or optimistic) RUE flow pattern, leading to a worse system performance than what the system manager expected. To avoid such issue, the risk-averse system manager minimizes the objective function by selecting an optimal capacity expansion vector and the corresponding *worst-case RUE path flow pattern* (i.e., the RUE path flow that yields the largest objective function value). This is achieved by formulating the design problem as a min-max optimization problem. In summary, the *capacity expansion RUE-NDP under cost recovery* is formulated as

$$\min_{\mathbf{y}\in\Omega_y}\max_{\mathbf{f}\in\Omega_f}\left(TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}(\mathbf{f}),\mathbf{y})+IC(\mathbf{y})\right), \tag{15}$$

subject to (11), (12), (13), and (14).

The proposed problem is a bi-level optimization problem with equilibrium constraints. The bi-level optimization problem refers to the min-max problem (15). The first level (lower level) problem is to find the worst RUE path flow pattern and its corresponding minimum generalized path travel cost budget vector that yield the maximum objective function value for a given capacity expansion vector. The second level (upper level) problem is to minimize the maximum objective function value by selecting an optimal capacity expansion vector. The equilibrium constraints are presented by the system of non-linear equalities and inequalities (11)-(12).

The objective function is continuous and differentiable in terms of link flows and link capacities. The feasible solution set is non-empty and compact. Therefore, an optimal solution to the bi-level optimization problem with equilibrium constraints, denoted as $(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})$, must exist. Similar to other network design problems, the upper level problem can be solved by many heuristics such as Genetic Algorithm. The lower level problem can be solved by all-or-nothing assignment. It is well-known that optimal solutions to a bi-level optimization problem may not be unique. However, the minima of the objective function must be unique.

For the ease of presentation, we use *Problem Q* to refer to the proposed min-max capacity expansion RUE-NDP under cost recovery. We examine the PoA of Problem Q in the following section.

4 Analysis on the PoA

Problem Q is a member of the family of RUE-NDPs formulated in Sub-section 2.1. The PoA for Problem Q, $\rho\left(G,\mathbf{d},\mathbf{t},\theta\right)$, follows its definition in (8), where θ is $\left(\boldsymbol{\sigma},R^{\mathrm{t}},R^{\mathrm{s}},R^{\mathrm{u}}\right)$, $Z(\mathbf{v}(\mathbf{f}),\mathbf{y})$ is $TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}(\mathbf{f}),\mathbf{y})+IC(\mathbf{y})$, the equilibrium solution $(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}})$ is $(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})$. The system optimal solution, denoted by $(\mathbf{v}^{*}(\mathbf{f}^{*}),\mathbf{y}^{*})$, is obtained by solving the following RSO-NDP:

$$\min_{\mathbf{y}\in\Omega_y,\ \mathbf{f}\in\Omega_f}\left(TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}(\mathbf{f}),\mathbf{y})+IC(\mathbf{y})\right). \tag{16}$$

The solution $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)$ must exist because of the following reasons: 1) the TSTCB and the IC are continuous functions in terms of path flows and link capacities; 2) the solution set is non-empty and compact. Because the objective function in (16) is non-convex, the optimal solutions $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)$ are non-unique. Nevertheless, the minimum objective function value must be unique.

We present a novel approach to deriving the analytical formulation of an upper bound of $\rho\left(G,\mathbf{d},\mathbf{t},\theta\right)$.

4.1 Properties of the equilibrium and system optimal solutions

Prior to the analysis of the properties, the following parameter is introduced. Denote $\varepsilon_{\max}$ as the *maximum ratio between link travel time standard deviation and mean link travel time*, i.e., $\varepsilon_{\max}=\max_{a\in A}\left(\sigma_a/t_a\right)$. The parameter $\varepsilon_{\max}$ must exist because the link travel time standard deviations and the mean link travel times of all links are finite. The value of $\varepsilon_{\max}$ can be theoretically derived or calibrated from travel time data.

Given $\mathbf{y}^{RUE}$, we prove the following:

Property 1. Given $\mathbf{y}^{RUE}$, let $\mathbf{f}'''=(f_p''')_{p\in P}$ and $\mathbf{v}'''(\mathbf{f}''')$ be the path flow pattern and the corresponding link flow pattern that minimizes the sum of individual path travel cost budgets. The ratio between the sum of individual path travel cost budgets of an RUE flow pattern $\mathbf{f}^{RUE}=(f_p^{RUE})_{p\in P}$ and that of the flow pattern $\mathbf{f}'''$ is bounded above:

$$\sum_{p\in P} f_p^{RUE} b_p^{RUE} \Big/ \sum_{p\in P} f_p''' \, b_p''' \le 1+\varepsilon_{\max} R^{\mathrm{u}} / R^{\mathrm{t}} \, ,$$

where $b_p^{RUE}=b_p(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})$ and $b_p'''=b_p(\mathbf{v}'''(\mathbf{f}'''),\mathbf{y}^{RUE})$, $\forall p\in P$.

Proof. See Appendix C.

Property 2. Given $\mathbf{y}^{RUE}$, let $\mathbf{v}''(\mathbf{f}'')$ be the corresponding link flow pattern of the path flow pattern $\mathbf{f}''$ that minimizes the TSTCB. The ratio between the TSTCB of an RUE link flow pattern $\mathbf{v}^{RUE}(\mathbf{f}^{RUE})$ and that of the flow pattern $\mathbf{v}''(\mathbf{f}'')$ is bounded above:

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE}) \Big/ TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}''(\mathbf{f}''),\mathbf{y}^{RUE}) \le \left(1+\varepsilon_{\max} R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max} R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2 .$$

Proof. See Appendix D.

Property 2 can be interpreted as follows: The *inefficiency of the worst RUE flow pattern* given $\mathbf{y}^{RUE}$ with respect to the system performance measure is bounded above.

Given $\mathbf{y}^*$, we further prove the following:

Property 3. Given $\mathbf{y}^*$, assume $\bar{\mathbf{f}}$ and $\bar{\mathbf{v}}(\bar{\mathbf{f}})$ are the worst RUE path flow and link flow patterns yielding the largest objective function value, respectively. The ratio between the TSTCB of $\bar{\mathbf{v}}(\bar{\mathbf{f}})$ and the TSTCB of $\mathbf{v}^*(\mathbf{f}^*)$ is bounded above:

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)\Big/TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)\le\left(1+\varepsilon_{\max}R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max}R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2.$$

Proof. This is a direct result of Property 2 if 1) $\mathbf{y}^{RUE}$ is replaced by $\mathbf{y}^*$; 2) $\mathbf{v}^{RUE}(\mathbf{f}^{RUE})$ is replaced by $\bar{\mathbf{v}}(\bar{\mathbf{f}})$; and 3) $\mathbf{v}''(\mathbf{f}'')$ is replaced by $\mathbf{v}^*(\mathbf{f}^*)$.

Property 4. Given $\mathbf{y}^*$, the ratio between the objective function value of $\bar{\mathbf{v}}(\bar{\mathbf{f}})$ defined in Property 1 and that of $\mathbf{v}^*(\mathbf{f}^*)$ is bounded above:

$$\begin{aligned}&\left(TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)+IC(\mathbf{y}^*)\right)\Big/\left(TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)+IC(\mathbf{y}^*)\right)\\&\le\left(1+\varepsilon_{\max}R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max}R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2.\end{aligned}$$

Proof. The following is true: Given three positive numbers g_1, g_2, and g_3. If g_1 is larger than or equal to g_2, then $(g_1+g_3)/(g_2+g_3)\le g_1/g_2$. Replacing g_1 with $TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)$, g_2 with $TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)$, g_3 with $IC(\mathbf{y}^*)$, and using Property 3, the result is obtained. ■

4.2 Upper bound of the PoA and its properties

Based on Property 4, we prove that an upper bound of the PoA exists as shown below.

Proposition 1. Given an instance $(G,\mathbf{d},\mathbf{t},\theta)$, the price of anarchy $\rho(G,\mathbf{d},\mathbf{t},\theta)$ is bounded above:

$$\rho(G,\mathbf{d},\mathbf{t},\theta)\le\left(1+\varepsilon_{\max}R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max}R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2. \qquad (17)$$

Proof. The solution $(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)$ is a feasible solution, but it may not be the equilibrium solution because $\mathbf{y}^*$ may not be $\mathbf{y}^{RUE}$. Thus, the objective function value of $(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})$ is not larger than that of $(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)$, i.e.,

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})+IC(\mathbf{y}^{RUE})\le TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)+IC(\mathbf{y}^*).$$

Dividing both sides of the above inequality by the objective function value of $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)$, the following inequality is obtained:

$$\frac{TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})+IC(\mathbf{y}^{RUE})}{TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{*}(\mathbf{f}^{*}),\mathbf{y}^{*})+IC(\mathbf{y}^{*})}\le\frac{TSTCB_{R\,,R^{\mathrm{t}}\,\mathrm{s}}(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^{*})+IC(\mathbf{y}^{*})}{TSTCB_{R\,,R^{\mathrm{t}}\,\mathrm{s}}(\mathbf{v}^{*}(\mathbf{f}^{*}),\mathbf{y}^{*})+IC(\mathbf{y}^{*})}. \quad (18)$$

The left side of (18) is precisely the price of anarchy $\rho(G,\mathbf{d},\mathbf{t},\theta)$. According to Property 4, the right side of inequality (18) is not larger than $\left(1+\varepsilon_{\max}R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max}R^{\mathrm{u}}/R^{\mathrm{t}}\right)^{2}$. It means that the left side of inequality (18), which is $\rho(G,\mathbf{d},\mathbf{t},\theta)$, is also bounded above by $\left(1+\varepsilon_{\max}R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max}R^{\mathrm{u}}/R^{\mathrm{t}}\right)^{2}$.■

The derived upper bound of the PoA is dependent on $\varepsilon_{\max}$, R^{t}, R^{u}, and R^{s}, which are the maximum ratio between link travel time standard deviation and mean link travel time, the VOT, the VOR for path travel time, and the VOR for system travel time, respectively. The sensitivities of the upper bound of the PoA with respect to these parameters are addressed in the following.

Property 5. The upper bound of the PoA is increasing with respect to $\varepsilon_{\max}$, R^{u}, and R^{s}. The upper bound of the PoA is decreasing with respect to R^{t}.

The following figures present the sensitivities of upper bounds of PoAs subject to parameters $\varepsilon_{\max}$, R^{u}, R^{s}, and R^{t}.

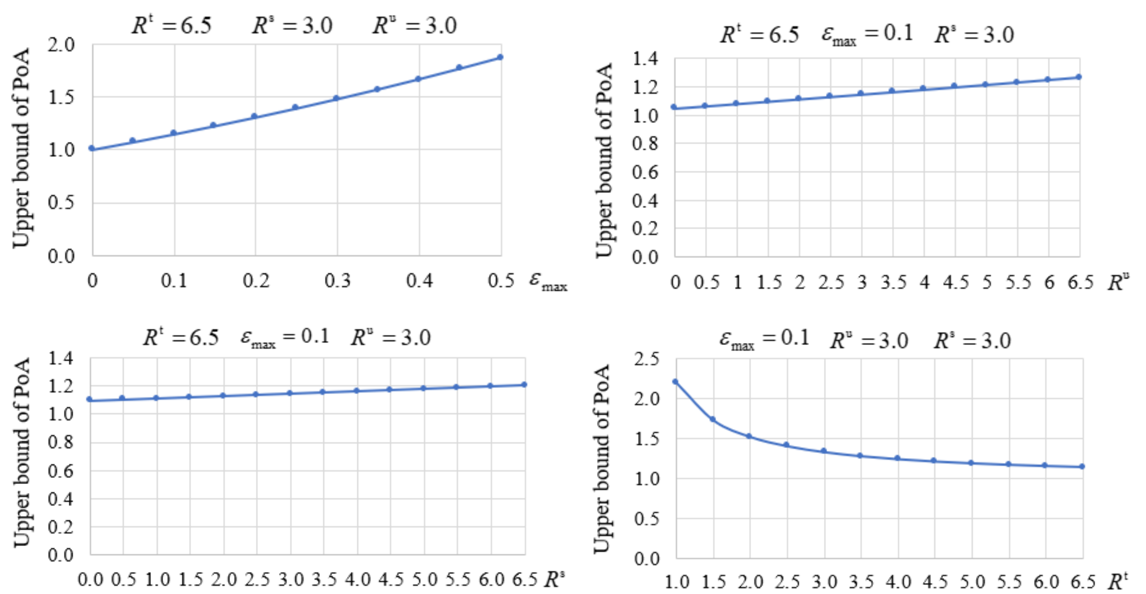


Figure 1. The upper bounds of PoAs given different parameter values

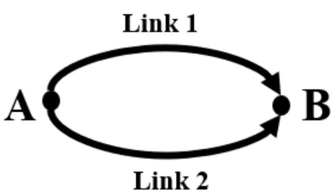
Property 6. The upper bound of the PoA for the capacity expansion RUE-NDP under cost recovery is independent of network topology and travel demands.

Property 7. The upper bound of the PoA for the capacity expansion RUE-NDP under cost recovery is independent of travel time functions.

The proofs of Properties 5, 6, and 7 are straightforward and omitted.

In the following, we present an example to illustrate how to calculate the upper bound of PoA given a design instance.

Example 1:



Design instance:

$R^{t} = 6.0 \quad R^{u} = 2.3 \quad R^{s} = 2.3$

$d_{AB} = 5.0 \quad \kappa_1 = 2.0 \quad \kappa_2 = 3.0$

$t_1 = 2.0 + 0.5x_1/(y_1+2)$

$t_2 = 2.2 + 0.3x_2/(y_2+3)$

$\sigma_1 = 0.2 + 0.01x_1/(y_1+2)$

$\sigma_2 = 0.1 + 0.03x_2/(y_2+3)$

Solution:

$x_1^{RUE} = 2.04 \quad y_1^{*} = 0.19$

$x_2^{RUE} = 2.96 \quad y_2^{*} = 0$

$t_1^{RUE} = 2.47 \quad \sigma_1^{RUE} = 0.21$

$t_2^{RUE} = 2.50 \quad \sigma_2^{RUE} = 0.13$

To get an upper bound of the PoA for this design instance, we need the parameters R^{s}, R^{t}, R^{u}, and $\varepsilon_{\max}$. $\varepsilon_1 = 0.21/2.47 = 0.08$ and $\varepsilon_2 = 0.13/2.50 = 0.05$. Take $\varepsilon_{\max} = 0.08$. We also have $R^{t} = 6.0$, $R^{s} = 2.3$, and $R^{u} = 2.3$. The upper bound of PoA is $\left(1+\varepsilon_{\max}R^{s}/R^{t}\right)\left(1+\varepsilon_{\max}R^{u}/R^{t}\right)^2 = 1.10$.

In this example, the upper bound of PoA is independent of the network topology, travel demands, and travel time functions, as indicated in Property 6 and Property 7.

Based on Proposition 1 and Property 5, the following proposition can be directly concluded.

Proposition 2. Denote I as the set of instances in which each instance satisfies the following conditions: 1) the maximum ratio between link travel time standard deviation and mean link travel time does not exceed $\bar{\varepsilon}_{\max}$; 2) the VOR for path travel time does not exceed $\bar{R}^{u}$; 3) the VOR for system travel time does not exceed $\bar{R}^{s}$; and 4) the VOT is not less than $\bar{R}^{t}$. The price of anarchy of I is bounded above:

$$\rho(I) \le \left(1+\bar{\varepsilon}_{\max}\bar{R}^{s}/\bar{R}^{t}\right)\left(1+\bar{\varepsilon}_{\max}\bar{R}^{u}/\bar{R}^{t}\right)^2. \tag{19}$$

Remark 1. The existence of an upper bound indicates that both $\rho(G,\mathbf{d},\mathbf{t},\theta)$ and $\rho(I)$ for Problem Q are not trivial notions (i.e., the PoA is meaningless if it is unbounded).

Remark 2. The upper bound of the PoA equals one under either of the two conditions: 1) there is no supply uncertainty (i.e., $\bar{\varepsilon}_{\max}$ = 0); 2) there are no monetary values in the reliabilities of system travel time and path travel time (i.e., $\bar{R}^{\mathrm{s}} = \bar{R}^{\mathrm{u}} = 0$). The reason is that link marginal mean cost tolls are equivalent to link marginal cost tolls when there is no uncertainty, and charging the link marginal cost tolls drives the travelers to choose paths to minimize the TSTT. In Sub-section 2.2, it is discussed that the PoA for an instance set is intuitively larger than or equal to one. Together with the fact that the upper bound of the PoA is equal to one under either of the conditions, the PoA must equal one. Then, the upper bound of the PoA is equal to the PoA itself.

4.3 Discussions on the upper bound of the PoA

4.3.1 Application of the upper bound

The upper bound of the PoA carries different implications from those of $\rho(G,\mathbf{d},\mathbf{t},\theta)$ and $\rho(I)$. As mentioned in Sub-section 2.2, $\rho(G,\mathbf{d},\mathbf{t},\theta)$ reflects the exact inefficiency of the equilibrium solution given an instance $(G,\mathbf{d},\mathbf{t},\theta)$ and $\rho(I)$ reflects the exact worst-case inefficiency of the equilibrium solutions given a group of instances. The $\rho(G,\mathbf{d},\mathbf{t},\theta)$ and $\rho(I)$ are valuable economic evaluation indexes. The upper bound of the PoA, on the other hand, *is a quick estimate of* $\rho(G,\mathbf{d},\mathbf{t},\theta)$ or $\rho(I)$. Furthermore, computing the upper bound of the PoA only requires the values of a few parameters, which is an advantage when available information is limited. For example, computing $\rho(G,\mathbf{d},\mathbf{t},\theta)$ requires the information of G, $\mathbf{d}$, $\mathbf{t}$, and θ, where θ refers to the information related to link (additional) capacity and link free flow travel time variations. Acquiring this piece of information can be time and resource consuming. On the other hand, computing the upper bound of $\rho(G,\mathbf{d},\mathbf{t},\theta)$ only requires the information of $\varepsilon_{\max}$, R^{t}, R^{u}, and R^{s}, which can be acquired more easily. The system manager or other analysts can quickly and easily estimate the inefficiency of the equilibrium solution and decide if necessary measures are needed to deal with the selfish-routing behavior of travelers.

4.3.2 Comparison with existing studies

We proceed to compare the properties of the upper bound of the PoA for Problem Q to those of the upper bound of the PoA for the RUE-NDP proposed by Szeto and Wang (2015). Property 6 is consistent with the result of Szeto and Wang (2015).

Property 7, however, differs from the result of Szeto and Wang (2015), which indicates that the upper bound of the PoA is dependent on the highest degree of the mean link travel time functions. An implication of Property 7 is that the system manager does not need to acquire the information regarding travel time functions to calculate an *upper bound* of the PoA. Property 7 also implies that the derived upper bound is a bound of the PoAs for design instances in which the travel time functions can take any forms as long as they are differentiable and monotone increasing with respect to the link flows. The upper bound of the PoA proposed by Szeto and Wang (2015), on the other hand, can only bound the PoAs for design instances in which the travel time functions must be polynomial functions.

In the following, we explain why our proposed upper bound of the PoA has Property 7 and the upper bound of the PoA proposed by Szeto and Wang (2015) does not have. Szeto and Wang (2015) assumed specifically that the mean link travel time functions must be polynomial functions with respect to link flows. Our study only assumes that the mean link travel time functions are monotone increasing and differentiable functions with respect to link flows. Szeto and Wang (2015) used the mathematical properties of polynomial mean link travel time functions to derive an upper bound of the inefficiency of the RUE flow pattern, which is dependent on the highest degree of the mean link travel time functions. In our study, because the travelers are charged with link marginal mean cost tolls, we can derive an upper bound of the inefficiency of the worst RUE flow pattern given an optimal capacity expansion vector without knowing the explicit expressions of the mean link travel time functions. Thus, the upper bound of the PoA proposed by Szeto and Wang (2015) is dependent on the link travel time functions whereas ours is not.

5 Conclusion

The study proposed a general definition of the PoA for capacity expansion RUE-NDPs with the following features: 1) the objective function can include total system travel time, travel time reliability, construction cost, environmental cost, and other system performance measures; 2) auxiliary decision variables can be included as long as they do not affect the value of the objective function; 3) the lower level problem can be any type of RUE problems; and 4) additional constraints are incorporated.

This study proposed a capacity expansion RUE-NDP under cost recovery that considers supply uncertainty. The link marginal mean cost tolls are charged upon the travelers, and a cost recovery constraint is incorporated to guarantee that the degree of cost recovery (proposed and defined in this study) is larger than or equal to one. The problem is formulated as a min-max problem

A novel approach to deriving the analytical formula of an upper bound of the PoA is presented. The upper bound is independent of travel time functions, demands, and network topology. The upper bound is related to the travel time variations, the VORs for system travel time and path travel time, and the VOT. The upper bound is a quick estimate of the PoA value when limited information is available.

Acknowledgments

This work was jointly supported by a grant [No. 201711159034] from the University Research Committee of the University of Hong Kong, a grant from the National Natural Science Foundation of China [No. 71771194], and a grant from the Guangzhou Science Technology and Innovation Commission [No. 201707010292]. The authors are grateful to the Editor-in-Chief and two reviewers for their constructive comments.

References

Abdulaal M, LeBlanc LJ (1979) Continuous equilibrium network design models.

Transportation Research Part B 13(1): 19-32

Ben-Ayed O, Boyce DE, Blair III CE (1988) A general bilevel linear programming formulation of the network design problem. Transportation Research Part B 22(4): 311-318

Brownstone D, Small KA (2003) Valuing time and reliability: assessing the evidence from road pricing demonstrations. Transportation Research Part A 39(4): 279-293

Chau CK, Sim KM (2003) The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands. Operations Research Letters 31(5): 327-334

Chen A, Kim J, Zhou Z, Chootinan P (2007) Alpha reliable network design problem. Transportation Research Record 2029: 49-57

Chen A, Zhou Z (2010) The α-reliable mean-excess traffic equilibrium model with stochastic travel times. Transportation Research Part B 44(4): 493-513

Chen A, Zhou Z, Chootinan P, Ryu S, Yang C, Wong SC (2011) Transport network design problem under uncertainty: a review and new developments. Transport Reviews 31(6): 743-768

Chiou S (2005) Bilevel programming for the continuous transport network design problem. Transportation Research Part B 39(4): 361-383

Chootinan P, Wong SC, Chen A (2005) A reliability-based network design problem. Journal of Advanced Transportation 39(3): 247-270

Chow JYJ, Regan AC (2011) Network-based real option models. Transportation Research Part B 45(4): 682-695

Concas S, Kolpakov A (2009) Synthesis of research on value of time and value of reliability. http://www.nctr.usf.edu/pdf/77806.pdf

Correa J, Schulz A, Stier-Moses N (2004) Selfish routing in capacitated networks. Mathematics of Operations Research 29(4): 961-976

Dantzig GB, Maier SF, Harvey RP, Lansdowne ZF, Robinson DW (1979) Formulating and solving the network design problem by decomposition. Transportation Research Part B 13(1): 5-17

Farahani RZ, Miandoabchi E, Szeto WY, Rashidi H (2013) A review of urban transportation network design problems. European Journal of Operational Research 229(2): 281-302

Friesz TL, Anandalingam G, Mehta NJ, Nam K, Shah SJ, Tobin RL (1993) The multiobjective equilibrium network design problem revisited: a simulated annealing approach. European Journal of Operation Research 65(1): 44-57

Guo XL, Yang H, Liu TL (2010) Bounding the inefficiency of logit-based stochastic user equilibrium. European Journal of Operational Research 201(2): 463-469

Han D, Lo HK, Yang H (2008) On the price of anarchy for non-atomic congestion games under asymmetric cost maps and elastic demands. Computers & Mathematics with Applications 56(10): 2737-2743

Han D, Yang H (2008) The multi-class, multi-criterion traffic equilibrium and the efficiency of congestion pricing. Transportation Research Part E 44(5): 753-773

Huang HJ, Liu TL, Guo XL, Yang H (2011) Inefficiency of logit-based stochastic user equilibrium in a traffic network under ATIS. Networks and Spatial Economics 11(2):

255-269
Ji XF, Ban XG, Li MT, Zhang J, Ran B (2017) Non-expected route choice model under risk on stochastic traffic networks. Networks and Spatial Economics 17(3): 777-807
Khani A, Boyles SD (2015) An exact algorithm for the mean-standard deviation shortest path problem. Transportation Research Part B 81: 252-266
Koutsoupias E, Papadimitriou C (1999) Worst-case equilibria. In: Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS): Trier, Germany, Lecture Notes in Computer Science, vol. 1563, Springer, Berlin, 404-413
LeBlanc LJ, Boyce DE (1986) A bilevel programming algorithm for exact solution of the network design problem with user-optimal flows. Transportation Research Part B 20(3): 259-265
Liu ZY, Yi W, Wang SA, Chen J (2017) On the uniqueness of user equilibrium flow with speed limit. Networks and Spatial Economics 17(3): 763-775
Lo HK, Luo XW. Siu BWY (2006) Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. Transportation Research Part B 40(9): 792-806
Lo HK, Szeto WY (2009) Time-dependent transport network design under cost-recovery. Transportation Research Part B 43(1): 142-158
Meng Q, Yang H (2002) Benefit distribution and equity in road network design. Transportation Research Part B 36(1): 19-35
Ng MW, Waller ST (2009) Reliable system-optimal network design: convex mean-variance model with implicit chance constraints. Transportation Research Record 2090: 68-74
Roughgarden T (2005) Selfish routing and price of anarchy. MIT Press. Cambridge, MA
Roughgarden T, Tardos E (2002) How bad is selfish routing? Journal of the ACM 49(2): 236-259
Shao H, Lam WHK, Tam ML (2006) A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. Networks and Spatial Economics 6(3): 173-204
Small KA, Yan J (2001) The value of "value pricing" of roads: second-best pricing and product differentiation. Journal of Urban Economics 49: 310-336
Sumalee A, Luathep P, Lam WHK, Connors RD (2009) Transport network capacity evaluation and design under demand uncertainty. Transportation Research Record 2090: 17-28
Szeto WY, Jaber XQ, O'Mahony M (2010) Time-dependent discrete network design frameworks considering land use. Computer-Aided Civil and Infrastructure Engineering 25(6): 411-426
Szeto WY, Lo HK (2005) Strategies for road network design over time: robustness under uncertainty. Transportmetrica 1(1): 47-63
Szeto WY, Lo HK (2008) Time-dependent transport network improvement and tolling strategies. Transportation Research Part A 42(2): 376-391
Szeto WY, Wang B (2015) Price of anarchy for reliability-based traffic assignment and network design. Transportmetrica A: Transport Science 11(7): 603-635
Szeto WY, Wang B (2016) Reliable network design under supply uncertainty with probabilistic guarantees. Transportmetrica A: Transport Science 12(6): 504-532

Szeto WY, Wang Y, Wong SC (2014) The chemical reaction optimization approach to solving the environmentally sustainable network design problem. Computer-Aided Civil and Infrastructure Engineering 29(2): 140-158

Tilahun N, Levinson DM (2009) Value of time comparisons in the presence of unexpected delay. Travel Demand Management and Road User Pricing: Success, Failure and Feasibility, Wafaa Saleh & Gerd Sammer, eds., 173-184, Ashgate Publishers.

Wang B, Szeto WY (2018) Reliability-based user equilibrium in a transport network under the effects of speed limits and supply uncertainty. Applied Mathematical Modelling 56: 186-201

Wang CL, Doan XV, Chen B (2014) Price of anarchy for non-atomic congestion games with stochastic demands. Transportation Research Part B 70: 90-111

Wardrop JG (1952) Some theoretical aspects of road traffic research. ICE Proceedings: Engineering Divisions 1(3): 325-362

Wu X (2015) Study on mean-standard deviation shortest path problem in stochastic and time-dependent networks: A stochastic dominance based approach. Transportation Research Part B 80(9): 275-290

Wu X, Nie Y (2011) Modeling heterogeneous risk-taking behavior in route choice: a stochastic dominance approach. Transportation Research Part A 45(9): 896-915

Xiao F, Yang H, Han D (2007) Competition and efficiency of private toll roads. Transportation Research Part B 41(3): 292-308

Yang H (1997) Sensitivity analysis for the elastic-demand network equilibrium problem with applications. Transportation Research Part B 31(1): 55-70

Yang H, Bell MGH (1998) Models and algorithms for road network design: a review and some new developments. Transport Reviews 18(3): 257-278

Yang H, Meng Q (2002) A note on “highway pricing and capacity choice in a road network under a build-operate-transfer scheme”. Transportation Research Part A 36(7): 659-663

Yin YF, Madanat SM, Lu XY (2009) Robust improvement schemes for road networks under demand uncertainty. European Journal of Operational Research 198(2): 470-479

Zhao XM, Wan CH, Bi J (2018) Day-to-day assignment models and traffic dynamics under information provision. Networks and Spatial Economics, https://doi.org/10.1007/s11067-018-9386-1

Appendix A. Lemma 1 and its proof

Lemma 1. For any link flow pattern $\mathbf{v}' = (v'_a)_{a\in A}$, the following inequality holds:

$$\begin{aligned}&\sum_{a\in A}\Big(R^{\mathrm{t}} t_a(v_a^{RUE}, y_a^{RUE}) + \tau_a(v_a^{RUE}, y_a^{RUE}) - R^{\mathrm{t}} t_a(v'_a, y_a^{RUE})\Big)v'_a \\ &\le \sum_{a\in A}\tau_a(v_a^{RUE}, y_a^{RUE})\cdot v_a^{RUE},\end{aligned} \tag{A.1}$$

where v_a^{RUE} and y_a^{RUE} denote the entries of $\mathbf{v}^{RUE}(\mathbf{f}^{RUE})$ and $\mathbf{y}^{RUE}$, respectively.

Proof. For an individual link $a \in A$, consider the following maximization problem:

$$\min_{x_a \ge 0} \bar{Z}_a(x_a) = R^{\mathrm{t}} \left(t_a(v_a^{RUE}, y_a^{RUE}) + v_a^{RUE} dt_a(v_a^{RUE}, y_a^{RUE}) / dv_a - t_a(x_a, y_a^{RUE}) \right) x_a .$$

The first order derivative of $\bar{Z}_a(x_a)$ with respect to x_a is

$$d\bar{Z}_a(x_a)/dx_a =$$
$$R^{\mathrm{t}} \left(t_a(v_a^{RUE}, y_a^{RUE}) + v_a^{RUE} dt_a(v_a^{RUE}, y_a^{RUE}) / dv_a - t_a(x_a, y_a^{RUE}) - x_a dt_a(x_a, y_a^{RUE}) / dv_a \right).$$

Because of the properties of the link travel time function and the marginal link cost toll function, the following hold: $d\bar{Z}_a(x_a)/dx_a > 0$ for $0 \le x_a < v_a^{RUE}$; $d\bar{Z}_a(x_a)/dx_a = 0$ for $x_a = v_a^{RUE}$, and $d\bar{Z}_a(x_a)/dx_a < 0$ for $x_a > v_a^{RUE}$. The objective function $\bar{Z}_a(x_a)$ is strictly increasing on $\left[0, v_a^{RUE}\right]$ and strictly decreasing on $\left(v_a^{RUE}, +\infty\right)$. If v_a^{RUE} equals zero, $d\bar{Z}_a(x_a)/dx_a = 0$ at $x_a = 0$ and $d\bar{Z}_a(x_a)/dx_a < 0$ for $x_a > 0$. The function $\bar{Z}_a(x_a)$ is strictly decreasing on $\left[0, +\infty\right)$. The global maximum point x_a^* of the objective function exists and is unique, and satisfies the condition: $d\bar{Z}_a(x_a^*)/dx_a = 0$, i.e., $x_a^* = v_a^{RUE}$.

Substituting the global maximum point x_a^* into the objective function $\bar{Z}_a(x_a)$, the maxima of the objective function is

$$\begin{aligned} \bar{Z}_a(x_a^*) &= R^{\mathrm{t}} \left(t_a(v_a^{RUE}, y_a^{RUE}) + v_a^{RUE} dt_a(v_a^{RUE}, y_a^{RUE}) / dv_a - t_a(v_a^{RUE}, y_a^{RUE}) \right) v_a^{RUE} \\ &= R^{\mathrm{t}}\, v_a^{RUE} v_a^{RUE} \cdot dt_a(v_a^{RUE}, y_a^{RUE}) / dv_a = \tau_a(v_a^{RUE}, y_a^{RUE}) \cdot v_a^{RUE}. \end{aligned}$$

Thus, given a feasible link flow v_a', the following inequality holds:

$$\begin{aligned} \bar{Z}_a(v_a') &= R^{\mathrm{t}} \left(t_a(v_a^{RUE}, y_a^{RUE}) + v_a^{RUE} dt_a(v_a^{RUE}, y_a^{RUE}) / dv_a - t_a(v_a', y_a^{RUE}) \right) v_a' \\ &= \left(R^{\mathrm{t}} t_a(v_a^{RUE}, y_a^{RUE}) + \tau_a(v_a^{RUE}, y_a^{RUE}) - R^{\mathrm{t}} t_a(v_a', y_a^{RUE}) \right) v_a' \qquad \text{(A.2)} \\ &\le \tau_a(v_a^{RUE}, y_a^{RUE}) \cdot v_a^{RUE}. \end{aligned}$$

Condition (A.2) holds for any individual link in the road network. Summing up condition (A.2) over all links on a path, the result (A.1) in the lemma is obtained. ■

Appendix B. Upper bounds of TSTCB and sum of individual path travel cost budgets

Based on the formula relating the path and link travel time standard deviation, the path travel time standard deviation is smaller than or equal to the sum of link travel time standard deviations of links on that path, i.e., $\varsigma_p \le \sum_{a \in A} \sigma_a \delta_p^a$. Similarly, $\sigma\left[TSTT\right] \le \sum_{a \in A} \sigma_a v_a$. According to the definition of $\varepsilon_{\max}$, we have $\varsigma_p \le \sum_{a \in A} \varepsilon_{\max} t_a \delta_p^a$ and $\sigma\left[TSTT\right] \le \sum_{a \in A} \varepsilon_{\max} t_a v_a$.

Because $\sigma\left[TSTT\right] \le \sum_{a\in A} \varepsilon_{\max} t_a v_a$, it can easily be proved that the TSTCB has an upper bound, which is the mean TSTT multiplied by a number:

$$TSTCB_{R^{t},R^{s}}(\mathbf{v}(\mathbf{f}),\mathbf{y}) \le \left(R^{t} + \varepsilon_{\max} R^{s}\right) \sum_{a\in A} t_a(v_a, y_a)\cdot v_a \,. \tag{B.1}$$

Similar to the sum of individual path travel cost budgets, we have

$$\sum_{p\in P} f_p b_p(\mathbf{v}(\mathbf{f}),\mathbf{y}) \le \left(R^{t} + \varepsilon_{\max} R^{u}\right) \sum_{a\in A} t_a(v_a, y_a)\cdot v_a \,. \tag{B.2}$$

Note that $\sum_{p\in P} f_p b_p(\mathbf{v}(\mathbf{f}),\mathbf{y}) = R^{t} \sum_{a\in A} t_a(v_a, y_a)\cdot v_a + R^{u} \sum_{p\in P} f_p \cdot \varsigma_p$.

Appendix C. Proof of Property 1

Proof. Assume $\mathbf{v}'''(\mathbf{f}''') = (v_a''')_{a\in A}$ is the link flow pattern of the path flow pattern $\mathbf{f}''' = (f_p''')_{p\in P}$ that minimizes the sum of individual path travel cost budgets. Let ς_p^{RUE} and ς_p''' be the path travel time standard deviations of $\mathbf{f}^{RUE}$ and $\mathbf{f}'''$, respectively. Let $\tau_a^{RUE} = \tau_a(v_a^{RUE}, y_a^{RUE})$, $\tau_a''' = \tau_a(v_a''', y_a^{RUE})$, $t_a^{RUE} = t_a(v_a^{RUE}, y_a^{RUE})$, $t_a''' = t_a(v_a''', y_a^{RUE})$, $q_p^{RUE} = \sum_{a\in A} t_a^{RUE} \delta_p^a$, $q_p''' = \sum_{a\in A} t_a''' \delta_p^a$, $\tilde{b}_p^{RUE} = b_p^{RUE} + \sum_{a\in A} \tau_a^{RUE} \delta_p^a$, and $\tilde{b}_p''' = b_p''' + \sum_{a\in A} \tau_a''' \delta_p^a$.

Because $\mathbf{f}^{RUE}$ is the RUE path flow pattern, the following inequality holds: $\sum_{p\in P} \left(f_p''' - f_p^{RUE}\right) \tilde{b}_p^{RUE} \ge 0$ (for details, see the solution method in Sub-section 2.7 in the study of Szeto and Wang 2016), which is equivalent to $\sum_{p\in P} \tilde{b}_p^{RUE} f_p^{RUE} \le \sum_{p\in P} \tilde{b}_p^{RUE} f_p'''$. Subtracting $\sum_{p\in P} \tilde{b}_p''' f_p'''$ from both sides of the above inequality, we obtain $\sum_{p\in P} \tilde{b}_p^{RUE} f_p^{RUE} - \sum_{p\in P} \tilde{b}_p''' f_p''' \le \sum_{p\in P} \left(\tilde{b}_p^{RUE} - \tilde{b}_p'''\right) f_p'''$, which can be rewritten as

$$\begin{aligned}
&\sum_{p\in P} b_p^{RUE} f_p^{RUE} + \sum_{p\in P} \left(\sum_{a\in A} \tau_a^{RUE} \delta_p^a\right) f_p^{RUE} - \sum_{p\in P} b_p''' f_p''' - \sum_{p\in P} \left(\sum_{a\in A} \tau_a^{RUE} \delta_p^a\right) f_p''' \le \\
&R^{t} \sum_{p\in P} \left(q_p^{RUE} - q_p'''\right) f_p''' + R^{u} \sum_{p\in P} \left(\varsigma_p^{RUE} - \varsigma_p'''\right) f_p''' + \sum_{p\in P} \left(\sum_{a\in A} \tau_a^{RUE} \delta_p^a - \sum_{a\in A} \tau_a''' \delta_p^a\right) f_p''',
\end{aligned}$$

or

$$\begin{aligned}
&\sum_{p\in P} b_p^{RUE} f_p^{RUE} - \sum_{p\in P} b_p''' f_p''' \\
&\le \left[R^{t} \sum_{p\in P} \left(q_p^{RUE} - q_p'''\right) f_p''' + \sum_{p\in P} \left(\sum_{a\in A} \tau_a^{RUE} \delta_p^a\right) f_p''' - \sum_{p\in P} \left(\sum_{a\in A} \tau_a^{RUE} \delta_p^a\right) f_p^{RUE} \right] \\
&+ R^{u} \sum_{p\in P} \left(\varsigma_p^{RUE} - \varsigma_p'''\right) f_p'''.
\end{aligned} \tag{C.1}$$

For the term in the square bracket on the right side of (C.1), we have: $R^{\mathrm{t}}\sum_{p\in P}\left(q_p^{RUE}-q_p^{\prime\prime\prime}\right)f_p^{\prime\prime\prime}=R^{\mathrm{t}}\sum_{a\in A}\left(t_a^{RUE}-t_a^{\prime\prime\prime}\right)v_a^{\prime\prime\prime}$, $\sum_{p\in P}\left(\sum_{a\in A}\tau_a^{RUE}\delta_p^a\right)f_p^{\prime\prime\prime}=\sum_{a\in A}\tau_a^{RUE}v_a^{\prime\prime\prime}$, and $\sum_{p\in P}\left(\sum_{a\in A}\tau_a^{RUE}\delta_p^a\right)f_p^{RUE}=\sum_{a\in A}\tau_a^{RUE}v_a^{RUE}$. Thus, the term in the square bracket in (C.1) can be expressed in terms of link-based variables:

$$\left[R^{\mathrm{t}}\sum_{p\in P}\left(q_p^{RUE}-q_p^{\prime\prime\prime}\right)f_p^{\prime\prime\prime}+\sum_{p\in P}\left(\sum_{a\in A}\tau_a^{RUE}\delta_p^a\right)f_p^{\prime\prime\prime}-\sum_{p\in P}\left(\sum_{a\in A}\tau_a^{RUE}\delta_p^a\right)f_p^{RUE}\right]= \sum_{a\in A}\left(R^{\mathrm{t}}t_a^{RUE}-R^{\mathrm{t}}t_a^{\prime\prime\prime}+\tau_a^{RUE}\right)v_a^{\prime\prime\prime}-\sum_{a\in A}\tau_a^{RUE}v_a^{RUE}. \tag{C.2}$$

According to Lemma 1 in Appendix A, the first term on the right side of inequality (C.2) is smaller than or equal to $\sum_{a\in A}\tau_a^{RUE}v_a^{RUE}$. Thus, the right side of (C.2) is smaller than or equal to zero. Because the term in the square bracket in (C.1) is smaller than or equal to zero, the left side of (C.1) is smaller than or equal to the second term on the right side of (C.1):

$$\sum_{p\in P}b_p^{RUE}f_p^{RUE}-\sum_{p\in P}b_p^{\prime\prime\prime}f_p^{\prime\prime\prime}\le 0+R^{\mathrm{u}}\sum_{p\in P}\left(\varsigma_p^{RUE}-\varsigma_p^{\prime\prime\prime}\right)f_p^{\prime\prime\prime}. \tag{C.3}$$

It is assumed that the mapping $\boldsymbol{\varsigma}=\left(\varsigma_p\right)_{\forall p\in P}$ is monotone in terms of path flow $\mathbf{f}$. Thus, the following holds:

$$R^{\mathrm{u}}\sum_{p\in P}\left(\varsigma_p^{\prime\prime\prime}-\varsigma_p^{RUE}\right)\left(f_p^{RUE}-f_p^{\prime\prime\prime}\right)\le 0\text{, or equivalently,}$$

$$R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{\prime\prime\prime}f_p^{RUE}+R^{\mathrm{u}}\sum_{p\in P}\left(\varsigma_p^{RUE}-\varsigma_p^{\prime\prime\prime}\right)f_p^{\prime\prime\prime}\le R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{RUE}f_p^{RUE}.$$

Eliminating the non-negative term $R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{\prime\prime\prime}f_p^{RUE}$ from the above inequality, the inequality still holds, i.e.,

$$R^{\mathrm{u}}\sum_{p\in P}\left(\varsigma_p^{RUE}-\varsigma_p^{\prime\prime\prime}\right)f_p^{\prime\prime\prime}\le R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{RUE}f_p^{RUE}. \tag{C.4}$$

Based on inequalities (C.3) and (C.4), the following is true:

$$\sum_{p\in P}b_p^{RUE}f_p^{RUE}-\sum_{p\in P}b_p^{\prime\prime\prime}f_p^{\prime\prime\prime}\le R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{RUE}f_p^{RUE}. \tag{C.5}$$

Based on (B.2), the following inequality holds:

$$\sum_{p\in P}f_pb_p\le\left(R^{\mathrm{t}}+\varepsilon_{\max}R^{\mathrm{u}}\right)\cdot\frac{1}{R^{\mathrm{t}}}\cdot\left(\sum_{p\in P}f_pb_p-R^{\mathrm{u}}\sum_{p\in P}f_p\cdot\varsigma_p\right)\text{, or equivalently,}$$

$$R^{\mathrm{u}}\sum_{p\in P}f_p\cdot\varsigma_p\le\left(1-\frac{R^{\mathrm{t}}}{R^{\mathrm{t}}+\varepsilon_{\max}R^{\mathrm{u}}}\right)\sum_{p\in P}f_pb_p.$$

Based on the above, the following inequality holds:

$$R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{RUE} f_p^{RUE} \le \left(R^{\mathrm{u}}\varepsilon_{\max}\Big/\left(R^{\mathrm{t}}+R^{\mathrm{u}}\varepsilon_{\max}\right)\right)\sum_{p\in P} b_p^{RUE} f_p^{RUE} .$$

The right side of the above inequality is an upper bound of the right side of (C.5) and thus is an upper bound of the left side of (C.5), which gives

$$\sum_{p\in P} b_p^{RUE} f_p^{RUE} - \sum_{p\in P} b_p''' f_p''' \le \left(R^{\mathrm{u}}\varepsilon_{\max}\Big/\left(R^{\mathrm{t}}+R^{\mathrm{u}}\varepsilon_{\max}\right)\right)\sum_{p\in P} b_p^{RUE} f_p^{RUE} ,$$

which further gives

$$\sum_{p\in P} b_p^{RUE} f_p^{RUE} \Big/ \sum_{p\in P} b_p''' f_p''' \le 1\Big/\left(1-\left(R^{\mathrm{u}}\varepsilon_{\max}\Big/\left(R^{\mathrm{t}}+R^{\mathrm{u}}\varepsilon_{\max}\right)\right)\right) = 1+\left(R^{\mathrm{u}}\varepsilon_{\max}\big/R^{\mathrm{t}}\right).$$

This completes the proof. ■

Appendix D. **Proof of Property 2**

Proof. Assume $\mathbf{v}''(\mathbf{f}'') = \left(v_a''\right)_{a\in A}$ is the link flow pattern of the path flow pattern $\mathbf{f}'' = \left(f_p''\right)_{p\in P}$ that minimizes the TSTCB given $\mathbf{y}^{RUE}$. Let $b_p'' = b_p(\mathbf{v}''(\mathbf{f}''), \mathbf{y}^{RUE})$.

By definition, the sum of individual path travel cost budgets is larger than or equal to the monetary value of mean TSTT, i.e.,

$$R^{\mathrm{t}}\sum_{a\in A} t_a^{RUE} v_a^{RUE} \le \sum_{p\in P} b_p^{RUE} f_p^{RUE} .$$

Multiplying both sides of the above inequality by $\left(1+\varepsilon_{\max}R^{\mathrm{s}}\big/R^{\mathrm{t}}\right)$, the inequality still holds. That is,

$$\left(R^{\mathrm{t}}+\varepsilon_{\max}R^{\mathrm{s}}\right)\sum_{a\in A} t_a^{RUE} v_a^{RUE} \le \left(1+\varepsilon_{\max}R^{\mathrm{s}}\big/R^{\mathrm{t}}\right)\sum_{p\in P} b_p^{RUE} f_p^{RUE} .$$

The left side of above inequality is an upper bound of the TSTCB according to (B.1) in Appendix B. Thus, the right side of the above inequality is larger than or equal to $TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}), \mathbf{y}^{RUE})$, i.e.,

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}), \mathbf{y}^{RUE}) \le \left(1+\varepsilon_{\max}R^{\mathrm{s}}\big/R^{\mathrm{t}}\right)\sum_{p\in P} b_p^{RUE} f_p^{RUE} . \tag{D.1}$$

Similarly, the following inequality holds:

$$\left(R^{\mathrm{t}}+\varepsilon_{\max}R^{\mathrm{u}}\right)\sum_{a\in A} t_a'' v_a'' \le \left(1+\varepsilon_{\max}R^{\mathrm{u}}\big/R^{\mathrm{t}}\right) TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}''(\mathbf{f}''), \mathbf{y}^{RUE}).$$

The left side of the above inequality is an upper bound of the sum of individual path travel cost budgets according to (B.2) in Appendix B. Thus, the right side of the above inequality is larger than or equal to $\sum_{p\in P} b_p'' f_p''$, which further gives

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}''(\mathbf{f}''), \mathbf{y}^{RUE}) \ge \left(\sum_{p\in P} b_p'' f_p''\right)\Big/\left(1+\varepsilon_{\max}R^{\mathrm{u}}\big/R^{\mathrm{t}}\right). \tag{D.2}$$

Dividing the left side of (D.1) by the left side of (D.2), and dividing the right side of (D.1) by the right side of (D.2), we obtain the following inequality:

$$\frac{TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})}{TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}''(\mathbf{f}''),\mathbf{y}^{RUE})} \le \frac{\sum_{p\in P} b_p^{RUE} f_p^{RUE}}{\sum_{p\in P} b_p'' f_p''}\left(1+\varepsilon_{\max} R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max} R^{\mathrm{u}}/R^{\mathrm{t}}\right). \quad \text{(D.3)}$$

In (D.3), $\sum_{p\in P} b_p'' f_p''$ is larger than $\sum_{p\in P} b_p''' f_p'''$ defined in Property 1, because $\sum_{p\in P} b_p''' f_p'''$ is the minimum sum of individual path travel cost budgets given $\mathbf{y}^{RUE}$. Thus,

$$\sum_{p\in P} b_p^{RUE} f_p^{RUE} \Big/ \sum_{p\in P} b_p'' f_p'' \le \sum_{p\in P} b_p^{RUE} f_p^{RUE} \Big/ \sum_{p\in P} b_p''' f_p''' .$$

Together with Property 1, we obtain the following inequality:

$$\sum_{p\in P} b_p^{RUE} f_p^{RUE} \Big/ \sum_{p\in P} b_p'' f_p'' \le 1+\varepsilon_{\max} R^{\mathrm{u}}/R^{\mathrm{t}} . \quad \text{(D.4)}$$

Inequalities (D.3) and (D.4) indicate that the left side of (D.3) is smaller than or equal to $\left(1+\varepsilon_{\max} R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max} R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2$. This completes the proof. ■

Bounding the inefficiency of the reliability-based continuous network design problem under cost recovery

Anny B. Wang[1,2] and W.Y. Szeto[3&*]
[1]Guangzhou Municipal Engineering Design & Research Institute, Guangdong, P. R. China
[2]School of Civil Engineering and Transportation, South China University of Technology, Guangdong, P. R. China
[3]Department of Civil Engineering, The University of Hong Kong, Hong Kong, P. R. China
[&]The University of Hong Kong Shenzhen Institute of Research and Innovation, Shenzhen, P. R. China
[*] Phone: (852) 28578552, Fax: (852) 25595337, Email: ceszeto@hku.hk

Abstract

This study defines the price of anarchy for general reliability-based transport network design problems, which is an indicator of inefficiency that reveals how much the design objective value exceeds its theoretical minimum value due to the risk averse and selfish routing behavior of travelers. This study examines a new problem, which is a reliability-based continuous network design problem under cost recovery. In this problem, the variations of system travel time and path travel times, the risk attitudes of the system manager and travelers, congestion toll charges, capacity expansions, and cost recovery constraint are explicitly considered. The design problem is formulated as a min-max problem with the reliability-based user equilibrium constraint. It is proved that the price of anarchy for this problem is bounded above, and the upper bound is independent of travel time functions, demands, and network topology. The upper bound is related to the travel time variations, the value of reliability, and the value of time.

Keywords: Inefficiency, price of anarchy, transport network design problem, reliability-based user equilibrium

1 Introduction

The *price of anarchy* (PoA), which was first termed by Koutsoupias and Papadimitriou (1999), measures the inefficiency of the traffic assignment problem. It reveals how much the system performance measure would exceed its theoretical minimum value when travelers choose routes selfishly. The PoA for traffic assignment problems has received great research attention. Four major lines of research have arisen (Roughgarden and Tardos 2002; Chau and Sim 2003; Correa et al. 2004; Roughgarden 2005; Xiao et al. 2007; Han and Yang 2008; Han et al. 2008; Guo et al. 2010; Huang et al. 2011; Wang et al. 2014; Szeto and Wang 2015), which are based on four considerations: arc capacity constraints; demand and link travel time/cost functions; road pricing; and extensions of traditional user equilibrium principles and multiple user classes. The PoA for traffic assignment problems is well understood by scholars. However, the PoA for other problems, e.g., network design problems (NDPs), has rarely been studied.

The NDPs have broad definitions (Farahani et al. 2013). The most popular family of NDPs in the literature is the family of capacity expansion NDPs (Abdulaal and LeBlanc 1979; Dantzig et al. 1979; LeBlanc and Boyce 1986; Ben-Ayed et al. 1988; Friesz et al. 1993; Yang 1997; Yang and Bell 1998; Meng and Yang 2002; Chiou 2005; Szeto and Lo 2005; Szeto et al. 2010; Szeto et al. 2014), which optimizes the system performance measures of the road networks by determining the optimal capacity expansions (i.e., the additional capacities added to existing roads and/or the capacities of new roads) and the flow pattern (i.e., the traffic flow distribution in the road network). Some of these NDPs are also known as user equilibrium network design problems (UE-NDPs) because they capture the selfish routing behavior of travelers, which means that the flow pattern must satisfy the user equilibrium (UE) constraints. These NDPs also have one common feature—they assume that the travel demands and link capacities are deterministic.

In reality, there are uncertainties in the travel demands and road supplies due to day-to-day travel demand fluctuation, special events, bad weather, road accidents, road construction activities, etc. The demand and supply uncertainties lead to system travel time and path travel time variations, which cannot be ignored by the system manager and travelers. The reliability-based user equilibrium network design problems (RUE-NDPs) are developed based on the deterministic UE-NDPs by considering demand uncertainty and/or supply uncertainty. Chen et al. (2011) conducted a detailed review of the family of RUE-NDPs (Chootinan et al. 2005; Chen et al. 2007; Ng and Waller 2009; Sumalee et al. 2009; Yin et al. 2009; Chow and Regan 2011; Szeto and Wang 2016). Most existing studies focus on the modeling, solution methods, and applications of the capacity expansion RUE-NDPs. However, the PoA for the capacity expansion RUE-NDPs, which is an important indicator for evaluating how much the design objective function value exceeds its theoretical minimum value when travelers chose routes selfishly, has rarely been studied.

Szeto and Wang (2015) proposed the PoA for a capacity expansion RUE-NDP. Their study was the first attempt in the literature to examine the inefficiency of transport NDPs with capacity expansions. Szeto and Wang (2015) illustrated that the PoA for their proposed RUE-NDP reveals how much the system performance measure may exceed its corresponding theoretical minimum value due to the inefficient allocation of system resources (i.e., capacity expansions) and traffic flow, the latter of which is caused by the selfish routing behavior of travelers. They proved that the PoA has an upper bound, indicating that the inefficiency of the resource allocation of the network design is bounded above. The study of Szeto and Wang (2015) is far from complete. Firstly, they only considered one member of the capacity expansion RUE-NDP family. Their proposed PoA may not reflect the inefficiencies of resource allocations of the other RUE-NDPs that have different design objectives, decision variables, and constraints. Secondly, their study implicitly assumed that the RUE flow pattern is unique given the capacity expansions. Thirdly, most RUE-NDPs assume that the project cost does not exceed the available budget. However, the project cost can also be fully recovered by charging congestion tolls upon the travelers (Yang and Meng 2002; Lo and Szeto 2009). For RUE-NDPs that consider toll charges, the PoAs proposed by Szeto and Wang (2015) are not suitable. Thus, a general definition of the PoA for capacity expansion RUE-NDPs is required.

This study expresses the family of capacity expansion RUE-NDPs in a generalized model formulation and proposes a general definition of the PoA for the capacity expansion RUE-NDPs. This study then considers a specific problem, which is *a capacity expansion RUE-NDP* under cost-recovery that considers supply uncertainty and road tolls. The problem is formulated as a min-max problem. The *min*-level problem aims to minimize the largest total system travel cost budget (TSTCB) plus the project cost. The TSTCB is a variant of the total system travel time budget and consists of the monetary cost of mean total system travel time and an extra cost associated with system travel time reliability. The *max*-level problem aims to determine the worst-case flow pattern that gives the largest TSTCB plus the project cost. The self-routing behavior and risk attitudes of travelers are captured by the reliability-based user equilibrium (RUE) constraints. In addition, travelers are charged with congestion tolls, which are used to recover the project cost. To guarantee that the project is self-financing or even profitable, a cost recovery constraint is incorporated. Based on the proposed model, this study proposes a novel approach to derive the analytical formula for an upper bound of the PoA.

The contributions of this study are as follows:

- We propose a general definition of the PoA for capacity expansion RUE-NDPs to measure the inefficiency of the reliability-based transport NDPs with capacity expansion and cost recovery;
- We propose a new NDP, namely capacity expansion RUE-NDP under cost recovery, in which the project cost is fully recovered by charging travelers with congestion tolls. It is formulated by a min-max approach; and
- It derives an analytical bound of the PoA of the proposed capacity expansion RUE-NDP under cost recovery.

The key findings regarding the upper bound of the PoA for the proposed RUE-NDP include the following:

- The upper bound depends on the travel time variations, the value of travel time, the value of reliability for system travel time, and the value of reliability for path travel time;
- The upper bound is independent of travel time functions, demands, and network topology; and
- The upper bound equals one if there are no travel time variations or/and the system manager and travelers are both risk-neutral, indicating that the PoA also equals one.

This paper is organized as follows. In Section 2, we express the family of capacity expansion RUE-NDPs in a generalized model formulation and propose a general definition of the PoA for the capacity expansion RUE-NDPs. In Section 3, we describe our new problem. In Section 4, we examine the PoA for the studied problem and evaluate its upper bound. In Section 5, we provide a concluding remark and discuss the future research directions.

2 PoA for the capacity expansion RUE-NDPs

Consider a road network with topology $G(N, A)$, in which N is a finite set of nodes and A is a finite set of directed links. The nodes represent existing or candidate intersections. The directed links represent roads whose existing capacities are to be expanded or whose capacities are to be determined. The network has multiple origin-destination (O-D) pairs that define where the travelers are from and where they head to. Each O-D pair is associated with its travel demand, which is the number of travelers between the origin and the destination per hour.

For the clarity of the presentation, the main notations are defined and introduced in Table 1.

Table 1. Notations

$\mathbb{R}$	The set of real numbers
$\mathbb{R}_+$	The set of positive real numbers
RS	The set of O-D pairs in the road network
P	The set of all possible paths connecting different O-D pairs in the road network; its size is denoted by $m \in \mathbb{R}_+$
P_{rs}	The set of all possible paths connecting O-D pair rs, $rs \in RS$
d_{rs}	The positive travel demand or mean travel demand between O-D pair $rs \in RS$
$\mathbf{d}$	The vector of travel demands/mean travel demands between all O-D pairs $(d_{rs})_{rs \in RS}$
δ_p^a	The link-path incidence indicator, which equals one if link $a \in A$ is on path $p \in P$, and equals zero otherwise
f_p	The non-negative flow or mean flow on path $p \in P$
$\mathbf{f}$	The vector of path flows or mean path flows $(f_p)_{p \in P}$
Ω_f	The set of feasible path flow patterns that satisfy the path-flow demand conservation constraints and non-negativity constraints: $\Omega_f = \left\{ \mathbf{f} \middle\| \sum_{p \in P_{rs}} f_p = d_{rs}, \forall rs \in RS;\ f_p \geq 0, \forall p \in P \right\}$
v_a	The non-negative flow or mean flow on link $a \in A$
$\mathbf{v}(\mathbf{f})$	The vector of link flows or mean link flows in the road network $(v_a)_{a \in A}$ with $v_a = \sum_{p \in P} f_p \delta_p^a$, $\forall a \in A$, $\mathbf{f} = (f_p)_{p \in P} \in \Omega_f$
y_a	The design variable, which is the capacity of a new link $a \in A$ or the link capacity expansion of an existing link $a \in A$
u_a	The upper bound of y_a, $a \in A$
Ω_y	The set of feasible link capacities or link capacity expansions: $\Omega_y = \{ \mathbf{y} \| 0 \leq y_a \leq u_a, \forall a \in A \}$

$\mathbf{y}$	The vector of the capacities of new links or link capacity expansions of existing links $\left(y_a\right)_{a\in A}$
$t_a(v_a, y_a)$	The mean link travel time function of link $a \in A$ in terms of its link flow and link capacity (expansion)
$\mathbf{t}$	The vector of mean link travel time functions $\left(t_a\right)_{a\in A}$
$\boldsymbol{\sigma}$	The covariance matrix, which contains all the link travel time variances and link travel time covariances

2.1 Generalized model formulation of capacity expansion RUE-NDPs

The capacity expansion RUE-NDPs have various input information known as the *design instances*. A design instance is described by the general form $\left(G, \mathbf{d}, \mathbf{t}, \theta\right)$, in which $\mathbf{d}$ and $\mathbf{t}$ are defined in Table 1, and θ stands for any additional and essential information related to the RUE-NDP. θ can be a scalar, a vector, or a set of vectors. For example, θ may include the project budget and the travel time variation related information.

A capacity expansion RUE-NDP is formulated as a bi-level mathematical optimization problem with decision variables, constraints, and an objective function.

The decision variables include the vector of capacity expansions (i.e., $\mathbf{y}$). The capacity expansions include the additional capacities added to existing roads and/or the capacities of new roads. Other decision variables include the path flow pattern $\mathbf{f}$. Note that the link flow pattern $\mathbf{v}(\mathbf{f})$ is dependent on the path flow pattern $\mathbf{f}$. Thus, the link flows are dependent variables. In an RUE-NDP, $\mathbf{y}$ and/or $\mathbf{f}$ may be random variables. The decision variables in the RUE-NDP are commonly the *mean* capacity expansions and *mean* link flows. In addition, in some NDPs (e.g., Szeto and Lo 2005, Lo and Szeto 2009), the travelers are charged with road tolls. The link tolls are commonly dependent variables whose values depend on the link flows. For convenience, we denote any auxiliary decision variables as a vector $\mathbf{w}$ whose feasible set is described by a non-empty set X_0.

The constraints of a capacity expansion RUE-NDP include the *feasibility constraints*, i.e., the path flow-demand conservation constraints, the link-path flow conservation constraints, the non-negativity constraints of path flows and capacity expansions, and the feasibility constraints of the auxiliary decision variables. These constraints are implicitly captured by the non-empty sets Ω_f, Ω_y, X_0, and the definition of $\mathbf{v}(\mathbf{f})$. Specifically, the constraint set Ω_y restricts which links can have capacity changed and which new links can be added, and hence any strategy would be embodied in this constraint set and in other additional constraints. Most importantly, the RUE-NDP incorporates a set of non-linear inequalities and equalities known as the *RUE constraints*. The constraints capture the self-routing behavior of travelers or the risk attitudes of travelers. Apart from the feasibility constraints and the RUE constraints, the RUE-NDP might also have other related constraints, such as the *budget constraint*, which guarantees that the project cost is not larger than the project budget. If travelers are charged with road tolls, the budget constraint can be replaced by the *cost recovery constraint* (Lo and Szeto 2009), which guarantees that the project cost is not larger than the total toll revenue collected from the travelers.

The *system performance measures* of the road network include the consumer surplus (Yang 1997), the reserve capacity (Yang and Bell 1998), the total vehicle miles (Friesz et al. 1993), the sum of total system travel time and construction cost (Chiou 2005), and the total system travel time/cost (Meng and Yang 2002). The objective function of an RUE-NDP includes the mean system performance measure (Chow and Regan 2011), the sum of the mean and (weighted) variance/standard deviation of the system performance measure (Ng and Waller 2009; Sumalee et al. 2009; Szeto and Wang 2016), and the worst-case value of the system performance measure (Yin et al. 2009). The objective function of the RUE-NDP is commonly a continuous function in terms of the decision variables, denoted as $Z(\cdot)$. In this study, we assume that the objective function value is dependent on the link flow pattern (or path flow pattern) and the capacity expansions, and is independent of the auxiliary decision variables.

For most existing capacity expansion RUE-NDPs, the objective is to minimize the objective function. However, such a design objective is optimistic when there are multiple link flow patterns for a given $\mathbf{y}$ (e.g., Liu et al., 2017). In fact, Wang and Szeto (2018) proved that the RUE link flow pattern is unique when two conditions hold: 1) the path travel costs are monotone in terms of path flows; 2) the link travel cost is a bijective function of link flow. If the RUE link flow pattern is non-unique, the actual RUE flow pattern after the implementation of the capacity expansions may be different from the design RUE flow pattern, yielding a worse system performance than what the system manager expected. To deal with this practical issue, we consider that the system manager (or a risk-averse system manager) aims to minimize the *worst* possible value of the objective function over $\mathbf{y}$, i.e., minimizing the maximum value of the objective function over $\mathbf{y}$.

Based on the above, we express the capacity expansion RUE-NDPs as the following *general* non-linear constrained optimization problem:

$$\min_{\mathbf{y}} \max_{\mathbf{f}} \; Z(\mathbf{v}(\mathbf{f}), \mathbf{y}), \tag{1}$$

subject to the *RUE constraints*:

$$\bar{g}_i(\mathbf{v}(\mathbf{f}), \mathbf{y}, \mathbf{w}) \le 0, \; i = 1, 2..., m \in \mathbb{R}^+, \tag{2}$$

$$g_i(\mathbf{v}(\mathbf{f}), \mathbf{y}, \mathbf{w}) = 0, \; i = 1, 2..., m \in \mathbb{R}^+; \tag{3}$$

the *feasibility constraints*:

$$\mathbf{f} \in \Omega_f, \mathbf{y} \in \Omega_y, \tag{4}$$

$$\mathbf{w} \in X_0; \tag{5}$$

and *other relevant sets of constraints (e.g., budget constraints or cost recovery constraints)*:

$$h_i(\mathbf{v}(\mathbf{f}), \mathbf{y}, \mathbf{w}) \le 0, \; i = 1, 2..., n \in \mathbb{R}_+, \tag{6}$$

where $\bar{g}_i$, g_i, and h_i are all functions of $\mathbf{v}(\mathbf{f})$, $\mathbf{y}$, and $\mathbf{w}$. In constraints (2) and (6), m is the total number of paths and n is the total number of additional constraints.

If the objective function $Z(\mathbf{v}(\mathbf{f}),\mathbf{y})$ is continuous, the set described by constraints (4), (5), and (6) is non-empty, and an RUE link flow pattern exists and satisfies the equilibrium constraints (2)-(3), then the optimization problem (1)-(6) has at least one optimal solution, denoted as $(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}},\tilde{\mathbf{w}})$. For any $\mathbf{y}$, if the RUE link flow pattern is unique, the problem (1)-(6) is equivalent to $\min_{\mathbf{f},\mathbf{y}} Z(\mathbf{v}(\mathbf{f}),\mathbf{y})$ subject to (2)-(6).

Remark. The generalized model formulation (1)-(6) can also be used to express the UE-NDPs, because the UE-NDPs are special cases of RUE-NDPs in which the travel time variations are zero and/or the travelers and the system manager are both risk-neutral.

2.2 General definition of the PoA for capacity expansion RUE-NDPs

Firstly, to show the rationality of defining the PoA for capacity expansion RUE-NDPs, we quote the statement of Roughgarden (2005): "The price of anarchy can be defined much more generally; indeed, the concept makes sense for every application possessing an objective function and a notion of equilibrium".

Secondly, we identify the theoretical minimum objective function value when all the travelers willingly choose paths to minimize the objective function value. The minimum objective function value is obtained by minimizing $Z(\mathbf{v}(\mathbf{f}),\mathbf{y})$ subject to the feasibility constraints (4), (5) and (6). The problem is referred to as a capacity expansion Reliability-based System Optimum NDP (RSO-NDP) and it is expressed as the following general non-linear minimization problem:

$$\min_{\mathbf{f}\in\Omega_f,\mathbf{y}\in\Omega_y} Z(\mathbf{v}(\mathbf{f}),\mathbf{y}). \tag{7}$$

The solution which yields the minimum objective function value is called the *system optimal solution*, and we denote it as $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*,\mathbf{w}^*)$. To differentiate the system optimal solution and the optimal solution to the RUE-NDP (i.e., $(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}},\tilde{\mathbf{w}})$), we call the latter the *equilibrium solution*.

Conceptually, the PoA is the worst-possible ratio between the objective function value of an equilibrium solution and that of a system optimal solution. A formal mathematical definition is given as follows.

i. Given an instance $(G,\mathbf{d},\mathbf{t},\theta)$ admitting a system optimal solution $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*,\mathbf{w}^*)$ and an equilibrium solution $(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}},\tilde{\mathbf{w}})$, the *PoA of* $(G,\mathbf{d},\mathbf{t},\theta)$ is

$$\rho(G,\mathbf{d},\mathbf{t},\theta) = Z(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}}) \big/ Z(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*). \tag{8}$$

ii. Denote the set of design instances that have some common features as I, e.g., the set of instances whose travel time functions are all Bureau of Public Road type link performance functions. The *PoA of* I is

$$\rho(I) = \sup_{(G,\mathbf{d},\mathbf{t},\theta)\in I} \rho(G,\mathbf{d},\mathbf{t},\theta). \tag{9}$$

Remark. The mathematical definition of the PoA may take different forms. For example, the pioneer study (Roughgarden 2005) included the two terms $(G,\mathbf{d},\mathbf{t},\theta)$ and I in the definition of the PoA for the classical traffic assignment problem (see Definition 2.3.1 (a) and (b) in his study), whereas some studies omitted them. In this study, we take the study of Roughgarden (2005) as the reference and include the two terms in the definition of the PoA for capacity expansion RUE-NDPs.

The PoA reflects the inefficiency of equilibrium solutions to the RUE-NDPs. The inefficiency refers to two aspects, which are both caused by the selfish-routing behavior of travelers: 1) the traffic flow distribution is not the best; and 2) the allocation of resources (capacity expansion) is not the best. In practice, the PoA is an economic evaluation index, based on which the system manager can quickly determine the relative reduction of system performance induced by the selfish-routing behavior of travelers brings to the transport network design. The PoA is a ratio and it is intuitively larger than one. A smaller PoA value indicates that the efficiency loss is less, and vice versa.

The $\rho(G,\mathbf{d},\mathbf{t},\theta)$ in (8) reflects the *exact inefficiency* of an equilibrium solution to the RUE-NDP with instance $(G,\mathbf{d},\mathbf{t},\theta)$. The $\rho(I)$ in (9), on the other hand, reveals the worst-case inefficiency of equilibrium solutions to the RUE-NDP with instances that share some common feature.

The PoA for the capacity expansion RUE-NDPs proposed in this study differs from the PoAs proposed by Szeto and Wang (2015). The PoAs proposed by Szeto and Wang (2015) are defined for the RUE-NDP that must satisfy the following conditions: 1) the lower level reliability-based user equilibrium flow patterns must be unique; 2) the decision variables are merely link capacity additions; 3) the design objective functions are total system travel time and total system travel time budget; 4) the RUE-NDP only considers supply uncertainty; and 5) the reliability-based user equilibrium problem adopts the travel time budget approach (Shao et al. 2006). The PoA proposed in our study, on the other hand, is defined for RUE-NDPs that satisfy less restrictive conditions. Firstly, the RUE-NDPs may have additional decision variables such as the road tolls. It allows the system manager to evaluate the impacts of the additional decision variables on the inefficiency of resource allocation. Secondly, apart from the classic system performance measure, which is the cost of system travel time, the objective functions may also include the cost of travel time reliability, environmental cost, construction cost, etc. It allows the system manager to evaluate the inefficiency of resource allocation with respect to different additional considerations such as travel time uncertainty, environmental impacts, and project cost, etc. Thirdly, the RUE-NDPs may incorporate additional constraints (e.g., the cost recovery constraint), which allows the system manager to evaluate the inefficiency of resource allocation when there are additional constraints to consider. Fourthly, the RUE-NDP may consider demand uncertainty/supply uncertainty or both, allowing the system manager to evaluate the inefficiency of resource allocation when the demand and/or supply are random variables. Finally, the lower level RUE problem of the RUE-NDPs may be formulated by other approaches. It allows the system manager to consider different types of RUE problems such as the mean-excess travel time (Chen and Zhou 2010) RUE problem, the stochastic dominance RUE problem (Wu and Nie 2011), and the non-expected route choice problem (Ji et al. 2017), etc.

To further illustrate the PoA in detail, we consider a specific problem proposed in the following, which is a capacity expansion RUE-NDP under cost recovery. The problem determines the capacities of the new roads in a road network under supply uncertainty and is formulated as a min-max problem. The travelers are charged with congestion tolls after the road network is built and put into usage. The construction cost of the road network is fully recovered from toll charges.

3 Reliability-based capacity expansion NDP under cost recovery: Min-max formulation

3.1 Objective function

Consider that the system manager designs which roads are expanded and/or built. Moreover, the manager considers the effect of supply uncertainty in the network design: the *actual link capacities may degrade from their design values (Szeto and Wang, 2015, 2016; Zhao et al., 2018) and the actual link free flow travel times may deviate from their pre-assumed values derived from maximum allowed speeds (Szeto and Wang, 2015, 2016).* The demands and the link flows are deterministic. The travel time on a link $a \in A$ (denoted by T_a) is thus modeled as a random variable.

From the system manager's perspective, his/her primary design objective is to minimize the total system travel time (TSTT). The TSTT equals the sum of the travel times experienced by all travelers. Thus, the TSTT is a compound random variable. We denote it as $\widehat{TSTT}$, and it equals $\widehat{TSTT} = \sum_{a \in A} T_a v_a$.

The expectation and standard deviation of the compound random variable $\widehat{TSTT}$ can be obtained by the following operations:

$$E\left[\widehat{TSTT}\right] = E\left[\sum_{a \in A} T_a v_a\right] = \sum_{a \in A} E\left[T_a\right] v_a ,$$

$$\sigma\left[\widehat{TSTT}\right] = \sigma\left[\sum_{a \in A} T_a v_a\right] = \left(\sum_{a \in A} \sigma^2\left[T_a\right] v_a^2 + \sum_{a \in A} \sum_{a' \in A, a \neq a'} v_a v_{a'} Cov\left[T_a, T_{a'}\right]\right)^{1/2} .$$

Commonly, the mean link travel time $E\left[T_a\right]$ of link $a \in A$ is predicted by its link travel time function $t_a(v_a, y_a)$. We assume that $t_a(v_a, y_a)$ is a *bijective* function with respect to its link flow given the link (additional) capacity. The link travel time function is monotone increasing and differentiable with respect to v_a, and monotone decreasing and differentiable with respect to y_a. We also assume that the link travel time variance $\sigma^2\left[T_a\right]$ and the travel time covariances $Cov\left[T_a, T_{a'}\right]$, $a' \in A$, $a' \neq a$ are finite. The explicit functional forms of the travel time variances depend on the link travel time functions and the assumed distributions of link free flow travel times and random link capacities.

Szeto and Wang (2015, 2016) proposed the concept of *total system travel time budget*, which simultaneously captures the mean and variation of TSTT, and is defined as:

Total system travel time budget = mean total system travel time + safety margin .

However, the system performance measure with a time unit is less preferable in practice because the investment parties are more concerned with the project cost rather than the TSTT itself. The system manager should consider the concerns of these parties. However, the TSTT cannot be directly combined with the project cost. Similarly, the total system travel time budget is also not a suitable indicator because it cannot be directly combined with the project cost. Thus, a similar concept to the total system travel time budget— *the TSTCB*—is proposed:

Total system travel cost budget = monetary value of mean total system travel time + monetary value of system travel time reliability.

The monetary value of mean TSTT can be obtained by multiplying the mean TSTT by a positive coefficient representing the value of time (VOT) for mean travel time:

$$\text{monetary value of mean TSTT} = \text{VOT} \cdot \text{mean TSTT},$$

in which the VOT is obtained by calibration using the survey data. The VOTs of road networks in different areas (e.g., cities, country regions, or countries) are different. Relevant studies on the VOT include the studies of Small and Yan (2001), Brownstone and Small (2003), and Tilahun and Levinson (2009).

The VOR converts a measure of travel time reliability into the monetary value of travel time reliability. The monetary value of travel time reliability can be obtained by

$$\text{monetary value of travel time reliability} = \text{VOR} \cdot \left(\text{measure of travel time reliability}\right).$$

The measures of travel time reliability include the difference between the 90th and 50th percentile travel time, the standard deviation of travel time, the difference between the actual late arrival and the usual travel time, and the difference between the early/late arrival time and the preferred arrival time. Given different measures of travel time reliability, the VORs are different. In this study, the standard deviation of TSTT is adopted as the measure of travel time reliability and used in the TSTCB.

Mathematically, the *TSTCB* is defined as follows:

$$TSTCB_{R^{\text{t}},R^{\text{s}}} = R^{\text{t}}\sum_{a\in A} E\left[T_a\right]v_a + R^{\text{s}}\sqrt{\sum_{a\in A}\sigma^{2}\left[T_a\right]v_a^{2} + \sum_{a\in A}\sum_{a'\in A, a\neq a'} v_a v_{a'} Cov\left[T_a, T_{a'}\right]},$$

in which R^{t} is the VOT for mean TSTT and R^{s} is the VOR for total system travel time.

There are no references for R^{s}. The report by Concas and Kolpakov (2009) only summarized the VORs for path travel time obtained by different studies. Nevertheless, the statistical methods used to calibrate the VOR for path travel time in that studies can also be used to calibrate R^{s}. Similar to the fact that the VOR for path travel time is dependent on the risk aversion of the travelers, R^{s} is related to the risk-aversion of the system manager. A larger R^{s} indicates that the system manager is more risk averse, and vice versa. The R^{s} equals zero if the system manager is risk neutral or/and considers that there is no monetary value in the reliability of TSTT.

As discussed before, apart from optimizing the system performance measure, the project cost is also an important consideration for the system manager. To formulate it, the annual cost of a link $a \in A$, denoted as $I_a(y_a)$, is introduced:

$$I_a(y_a) = \kappa_a \cdot y_a,\ \kappa_a > 0,\ \forall a \in A,$$

where the constant κ_a represents the annual cost per unit of (additional) capacity of link a. The annual cost per unit of (additional) capacity of a link $a \in A$ (i.e., κ_a) captures two factors: the annualized construction cost per unit of (additional) capacity and the annual maintenance cost per unit of (additional) capacity. The definition of $I_a(y_a)$ is based on two assumptions: 1) There is a constant return to scale in road construction, and 2) the maintenance/operation cost per unit of (additional) capacity is constant. The project cost equals the annual overall costs associated with the construction and maintenance of the road network, and we call it the *investment cost* (IC), which is

$$IC(\mathbf{y}) = \sum_{a \in A} I_a(y_a).$$

From the system manager's perspective, the design objective is to minimize the sum of the TSTCB and IC, i.e.,

$$\min_{\mathbf{y} \in \Omega_y, \mathbf{f} \in \Omega_f} TSTCB_{R^t, R^s}(\mathbf{v}(\mathbf{f}), \mathbf{y}) + IC(\mathbf{y}). \tag{10}$$

Note that if IC is not considered, then the above optimization model belongs to the family of mean-standard deviation models (e.g., Lo et al., 2006; Khani and Boyles, 2015; Wu, 2015).

3.2 RUE constraints with link marginal mean cost tolls

The travelers' selfish-routing and risk-adverse behaviors are captured by the RUE constraints. The RUE constraints are developed from Wardrop's first principle (Wardrop 1952), which states that a traveler always chooses a path that minimizes his/her own travel time. The travel time of a path equals the sum of the link travel times of all links on that path. Because the link travel times are all random variables, the path travel time, denoted as Q_p, $p \in P$, is also a random variable and expressed as

$$Q_p = \sum_{a \in A} T_a \delta_p^a, \ \forall p \in P.$$

The mean path travel time $E\left[Q_p\right]$, denoted as q_p, is $q_p = \sum_{a \in A} t_a \delta_p^a, \ \forall p \in P$.

When faced with travel time uncertainties, travelers often depart early and reserve extra time for their trips to avoid late arrivals. The risk-averse behavior of travelers is well known and many approaches extended from Wardrop's principle have been proposed to capture it. Among them, the path travel time budget (TTB) approach (Lo et al. 2006) is frequently adopted. The TTB approach assumes that each traveler selects a path with the minimum path TTB. The TTB is commonly defined as the sum of the mean path travel time and the weighted path travel time standard deviation.

Similar to the total system travel time budget, the path TTB also has a time unit. A variant of the TTB is the path travel cost budget, which has a cost unit and is defined as follows.

Path travel cost budget = monetary value of mean path travel time +
monetary value of path travel time reliability.

Similar to the TSTCB, the monetary values of mean path travel time and path travel time reliability can be obtained by the following operations:

monetary value of mean path travel time $= \text{VOT} \cdot q_p$, and monetary value of path travel time reliability = VOR $\cdot$ (measure of path travel time reliability),

in which the measure of path travel time reliability is the path travel time standard deviation. Based on the above, the *path travel cost budget* b_p, $\forall p \in P$ is

$$b_p = R^{\mathrm{t}} \cdot q_p + R^{\mathrm{u}} \cdot \sigma\left[Q_p\right],$$

in which $R^{\mathrm{t}} > 0$ is the VOT for mean path travel time and $R^{\mathrm{u}} \geq 0$ is the VOR for *path travel time.*

The VOT for mean path travel time and the VOT for mean total system travel time are consistent with each other, which are both R^{t}. As the measure of path travel time reliability is the path travel time standard deviation, the values for R^{u} can be found in the study of Concas and Kolpakov (2009).

It is assumed that all travelers are charged with congestion tolls because congestion toll charging has been adopting to mitigate congestion and improve system performance in reality. For a road network without uncertainties, *link marginal cost tolling* is one of the well-known tolling strategies for driving a UE flow pattern towards a flow pattern that yields a better system performance (Yang and Meng, 2002), and it is defined as *the product of the link flow and the first-order derivative of the link travel time function with respect to the link flow, assuming that the value of time is one*. For a road under supply uncertainty, however, because of the travel time variations, it is unclear whether charging the corresponding link marginal cost tolls will lead to an improvement in *TSTCB*. It only improves the *mean* TSTT. Nevertheless, this study assumes that the system manager adopts the link marginal cost tolls called *link marginal mean cost tolls* in a road network under supply uncertainty. The *link marginal mean cost toll* on link *a* is denoted by τ_a and defined by

$$\tau_a = R^{\mathrm{t}} v_a \cdot dt_a(v_a, y_a)/dv_a,\ \forall a \in A.$$

For a traveler, the *generalized path travel cost budget*, denoted by $\tilde{b}_p$, $\forall p \in P$, is

$$\tilde{b}_p = b_p + \sum_{a \in A} \delta_p^a \tau_a(v_a, y_a).$$

It is assumed that the travelers acquire the expectations and variabilities of path travel times, the VOT for path travel time, the VOR for path travel time standard deviation, and the link marginal mean cost tolls based on their experiences and factor this piece of information into their route choice considerations in the form of a generalized path travel cost budget. All travelers select routes to minimize their generalized path travel cost budgets. The long-term equilibrium is reached only if the generalized path travel cost budgets of all used routes are not higher than those of unused routes. The RUE flow path pattern $\mathbf{f}^{RUE} = (f_p^{RUE})_{p \in P}$ and the corresponding link flow pattern $\mathbf{v}^{RUE}(\mathbf{f}^{RUE})$ must satisfy the following *RUE constraints*:

$$f_p^{RUE}\left(\tilde{b}_p(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}), \mathbf{y}^{RUE}) - w_{rs}\right) = 0,\ \forall p \in P_{rs},\ \forall rs \in RS, \tag{11}$$

$$\tilde{b}_p(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}), \mathbf{y}^{RUE}) - w_{rs} \geq 0,\ \forall p \in P_{rs},\ \forall rs \in RS, \tag{12}$$

where w_{rs} is the minimum generalized path travel cost budget for O-D pair $rs \in RS$, and $\mathbf{y}^{RUE}$ is an optimal capacity solution to be determined. Denote $\mathbf{w} = \left(w_{rs}\right)_{rs\in RS}$ and it is the vector of auxiliary decision variables that must be non-negative, i.e.,

$$\mathbf{w} \geq \mathbf{0}. \tag{13}$$

Denote the standard deviation of the path travel cost as ς_p, $p \in P$. Unlike the mean link travel times, the mathematical property of ς_p is not known until the explicit formulation of link travel time standard deviations and travel time covariances are known. Without the loss of generality, we assume that the mapping $\boldsymbol{\varsigma} = \left(\varsigma_p\right)_{p\in P}$ is monotone with respect to the path flow pattern $\mathbf{f}$. Then, the path travel cost budgets are monotone with respect to the path flows. In addition, the mean link travel times are bijective functions of link flows. Following the proofs of Wang and Szeto (2018), the minimum path travel cost budgets, the monetary values of mean link travel times, and the RUE link flow pattern at equilibrium are unique. The RUE path flow pattern, on the other hand, is non-unique.

The generalized path travel cost budget includes the link marginal mean cost tolls. One of the purposes of charging link marginal mean cost tolls upon the travelers is to recover the IC. To check whether the total toll revenue collected from travelers covers the IC or not, the concept of the degree of cost recovery is introduced and defined in the next section.

3.3 Cost recovery constraint

A notion, namely the *degree of cost recovery*, denoted by η_τ, is defined as

$$\eta_\tau = \left(\boldsymbol{\tau}^{\mathrm{T}} \cdot \mathbf{v}(\mathbf{f})\right) \Big/ \left(\boldsymbol{\kappa}^{\mathrm{T}} \cdot \mathbf{y}\right),$$

where $\boldsymbol{\kappa}$ is the vector of the annual costs per unit of (additional) capacity defined in Sub-section 3.1.

The ratio defined in the above has been mentioned and adopted by Szeto and Lo (2008). The *degree of cost recovery* is an important indicator showing how profitable a toll scheme is. The project is profitable if η_τ is larger than one. The project is cost-recovery if η_τ is larger than or equal to one. The project is self-financing if η_τ exactly equals one. If η_τ is smaller than one, the total revenue collected from travelers cannot cover the IC, which means that the toll scheme $\boldsymbol{\tau}$ is not satisfactory from an investment perspective.

To guarantee that at an optimal design, the IC is fully covered by the total toll revenue collected from travelers, a cost recovery constraint is incorporated into the design problem. That is, the degree of cost recovery must be larger than or equal to one:

$$\eta_\tau \geq 1. \tag{14}$$

3.4 Model formulation

One possible way to depict the capacity expansion RUE-NDP under cost recovery is that it minimizes the objective function (i.e., (10)) subject to the RUE constraints (i.e., (11) and (12)), the cost recovery constraint (i.e., (14)), and the feasibility constraints of the decision variables. However, for a given $\mathbf{y}$, there might be multiple RUE link flow patterns that satisfy the RUE constraints. The objective (10) naturally selects the solution that has the minimum objective function value. In practice, the actual RUE flow pattern may deviate from the design (or optimistic) RUE flow pattern, leading to a worse system performance than what the system manager expected. To avoid such issue, the risk-averse system manager minimizes the objective function by selecting an optimal capacity expansion vector and the corresponding *worst-case RUE path flow pattern* (i.e., the RUE path flow that yields the largest objective function value). This is achieved by formulating the design problem as a min-max optimization problem. In summary, the *capacity expansion RUE-NDP under cost recovery* is formulated as

$$\min_{\mathbf{y}\in\Omega_y}\max_{\mathbf{f}\in\Omega_f}\left(TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}(\mathbf{f}),\mathbf{y})+IC(\mathbf{y})\right), \tag{15}$$

subject to (11), (12), (13), and (14).

The proposed problem is a bi-level optimization problem with equilibrium constraints. The bi-level optimization problem refers to the min-max problem (15). The first level (lower level) problem is to find the worst RUE path flow pattern and its corresponding minimum generalized path travel cost budget vector that yield the maximum objective function value for a given capacity expansion vector. The second level (upper level) problem is to minimize the maximum objective function value by selecting an optimal capacity expansion vector. The equilibrium constraints are presented by the system of non-linear equalities and inequalities (11)-(12).

The objective function is continuous and differentiable in terms of link flows and link capacities. The feasible solution set is non-empty and compact. Therefore, an optimal solution to the bi-level optimization problem with equilibrium constraints, denoted as $(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})$, must exist. Similar to other network design problems, the upper level problem can be solved by many heuristics such as Genetic Algorithm. The lower level problem can be solved by all-or-nothing assignment. It is well-known that optimal solutions to a bi-level optimization problem may not be unique. However, the minima of the objective function must be unique.

For the ease of presentation, we use *Problem Q* to refer to the proposed min-max capacity expansion RUE-NDP under cost recovery. We examine the PoA of Problem Q in the following section.

4 Analysis on the PoA

Problem Q is a member of the family of RUE-NDPs formulated in Sub-section 2.1. The PoA for Problem Q, $\rho(G,\mathbf{d},\mathbf{t},\theta)$, follows its definition in (8), where θ is $(\boldsymbol{\sigma},R^{\mathrm{t}},R^{\mathrm{s}},R^{\mathrm{u}})$, $Z(\mathbf{v}(\mathbf{f}),\mathbf{y})$ is $TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}(\mathbf{f}),\mathbf{y})+IC(\mathbf{y})$, the equilibrium solution $(\tilde{\mathbf{v}}(\tilde{\mathbf{f}}),\tilde{\mathbf{y}})$ is $(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})$. The system optimal solution, denoted by $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)$, is obtained by solving the following RSO-NDP:

$$\min_{\mathbf{y}\in\Omega_y,\ \mathbf{f}\in\Omega_f}\left(TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}(\mathbf{f}),\mathbf{y})+IC(\mathbf{y})\right). \tag{16}$$

The solution $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)$ must exist because of the following reasons: 1) the TSTCB and the IC are continuous functions in terms of path flows and link capacities; 2) the solution set is non-empty and compact. Because the objective function in (16) is non-convex, the optimal solutions $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)$ are non-unique. Nevertheless, the minimum objective function value must be unique.

We present a novel approach to deriving the analytical formulation of an upper bound of $\rho(G,\mathbf{d},\mathbf{t},\theta)$.

4.1 Properties of the equilibrium and system optimal solutions

Prior to the analysis of the properties, the following parameter is introduced. Denote $\varepsilon_{\max}$ as the *maximum ratio between link travel time standard deviation and mean link travel time*, i.e., $\varepsilon_{\max} = \max_{a\in A}\left(\sigma_a/t_a\right)$. The parameter $\varepsilon_{\max}$ must exist because the link travel time standard deviations and the mean link travel times of all links are finite. The value of $\varepsilon_{\max}$ can be theoretically derived or calibrated from travel time data.

Given $\mathbf{y}^{RUE}$, we prove the following:

Property 1. Given $\mathbf{y}^{RUE}$, let $\mathbf{f}''' = (f_p''')_{p\in P}$ and $\mathbf{v}'''(\mathbf{f}''')$ be the path flow pattern and the corresponding link flow pattern that minimizes the sum of individual path travel cost budgets. The ratio between the sum of individual path travel cost budgets of an RUE flow pattern $\mathbf{f}^{RUE} = (f_p^{RUE})_{p\in P}$ and that of the flow pattern $\mathbf{f}'''$ is bounded above:

$$\sum_{p\in P} f_p^{RUE} b_p^{RUE} \Big/ \sum_{p\in P} f_p''' \, b_p''' \le 1+\varepsilon_{\max} R^{\mathrm{u}} \big/ R^{\mathrm{t}} ,$$

where $b_p^{RUE} = b_p(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})$ and $b_p''' = b_p(\mathbf{v}'''(\mathbf{f}'''),\mathbf{y}^{RUE})$, $\forall p\in P$.

Proof. See Appendix C.

Property 2. Given $\mathbf{y}^{RUE}$, let $\mathbf{v}''(\mathbf{f}'')$ be the corresponding link flow pattern of the path flow pattern $\mathbf{f}''$ that minimizes the TSTCB. The ratio between the TSTCB of an RUE link flow pattern $\mathbf{v}^{RUE}(\mathbf{f}^{RUE})$ and that of the flow pattern $\mathbf{v}''(\mathbf{f}'')$ is bounded above:

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE}) \Big/ TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}''(\mathbf{f}''),\mathbf{y}^{RUE}) \le \left(1+\varepsilon_{\max} R^{\mathrm{s}} \big/ R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max} R^{\mathrm{u}} \big/ R^{\mathrm{t}}\right)^2 .$$

Proof. See Appendix D.

Property 2 can be interpreted as follows: The *inefficiency of the worst RUE flow pattern* given $\mathbf{y}^{RUE}$ with respect to the system performance measure is bounded above.

Given $\mathbf{y}^*$, we further prove the following:

Property 3. Given $\mathbf{y}^*$, assume $\bar{\mathbf{f}}$ and $\bar{\mathbf{v}}(\bar{\mathbf{f}})$ are the worst RUE path flow and link flow patterns yielding the largest objective function value, respectively. The ratio between the TSTCB of $\bar{\mathbf{v}}(\bar{\mathbf{f}})$ and the TSTCB of $\mathbf{v}^*(\mathbf{f}^*)$ is bounded above:

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)\Big/TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)\le\left(1+\varepsilon_{\max}R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max}R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2.$$

Proof. This is a direct result of Property 2 if 1) $\mathbf{y}^{RUE}$ is replaced by $\mathbf{y}^*$; 2) $\mathbf{v}^{RUE}(\mathbf{f}^{RUE})$ is replaced by $\bar{\mathbf{v}}(\bar{\mathbf{f}})$; and 3) $\mathbf{v}''(\mathbf{f}'')$ is replaced by $\mathbf{v}^*(\mathbf{f}^*)$.

Property 4. Given $\mathbf{y}^*$, the ratio between the objective function value of $\bar{\mathbf{v}}(\bar{\mathbf{f}})$ defined in Property 1 and that of $\mathbf{v}^*(\mathbf{f}^*)$ is bounded above:

$$\begin{aligned}&\left(TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)+IC(\mathbf{y}^*)\right)\Big/\left(TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)+IC(\mathbf{y}^*)\right)\\&\le\left(1+\varepsilon_{\max}R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max}R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2.\end{aligned}$$

Proof. The following is true: Given three positive numbers g_1, g_2, and g_3. If g_1 is larger than or equal to g_2, then $(g_1+g_3)/(g_2+g_3)\le g_1/g_2$. Replacing g_1 with $TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)$, g_2 with $TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)$, g_3 with $IC(\mathbf{y}^*)$, and using Property 3, the result is obtained. ■

4.2 Upper bound of the PoA and its properties

Based on Property 4, we prove that an upper bound of the PoA exists as shown below.

Proposition 1. Given an instance $(G,\mathbf{d},\mathbf{t},\theta)$, the price of anarchy $\rho(G,\mathbf{d},\mathbf{t},\theta)$ is bounded above:

$$\rho(G,\mathbf{d},\mathbf{t},\theta)\le\left(1+\varepsilon_{\max}R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max}R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2. \tag{17}$$

Proof. The solution $(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)$ is a feasible solution, but it may not be the equilibrium solution because $\mathbf{y}^*$ may not be $\mathbf{y}^{RUE}$. Thus, the objective function value of $(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})$ is not larger than that of $(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)$, i.e.,

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})+IC(\mathbf{y}^{RUE})\le TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\bar{\mathbf{v}}(\bar{\mathbf{f}}),\mathbf{y}^*)+IC(\mathbf{y}^*).$$

Dividing both sides of the above inequality by the objective function value of $(\mathbf{v}^*(\mathbf{f}^*),\mathbf{y}^*)$, the following inequality is obtained:

$$\frac{TSTCB_{R^{t},R^{s}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})+IC(\mathbf{y}^{RUE})}{TSTCB_{R^{t},R^{s}}(\mathbf{v}^{*}(\mathbf{f}^{*}),\mathbf{y}^{*})+IC(\mathbf{y}^{*})}\leq\frac{TSTCB_{R^{t},R^{s}}(\overline{\mathbf{v}}(\overline{\mathbf{f}}),\mathbf{y}^{*})+IC(\mathbf{y}^{*})}{TSTCB_{R^{t},R^{s}}(\mathbf{v}^{*}(\mathbf{f}^{*}),\mathbf{y}^{*})+IC(\mathbf{y}^{*})}. \quad (18)$$

The left side of (18) is precisely the price of anarchy $\rho\left(G,\mathbf{d},\mathbf{t},\theta\right)$. According to Property 4, the right side of inequality (18) is not larger than $\left(1+\varepsilon_{\max}R^{s}/R^{t}\right)\left(1+\varepsilon_{\max}R^{u}/R^{t}\right)^{2}$. It means that the left side of inequality (18), which is $\rho\left(G,\mathbf{d},\mathbf{t},\theta\right)$, is also bounded above by $\left(1+\varepsilon_{\max}R^{s}/R^{t}\right)\left(1+\varepsilon_{\max}R^{u}/R^{t}\right)^{2}$.■

The derived upper bound of the PoA is dependent on $\varepsilon_{\max}$, R^{t}, R^{u}, and R^{s}, which are the maximum ratio between link travel time standard deviation and mean link travel time, the VOT, the VOR for path travel time, and the VOR for system travel time, respectively. The sensitivities of the upper bound of the PoA with respect to these parameters are addressed in the following.

Property 5. The upper bound of the PoA is increasing with respect to $\varepsilon_{\max}$, R^{u}, and R^{s}. The upper bound of the PoA is decreasing with respect to R^{t}.

The following figures present the sensitivities of upper bounds of PoAs subject to parameters $\varepsilon_{\max}$, R^{u}, R^{s}, and R^{t}.

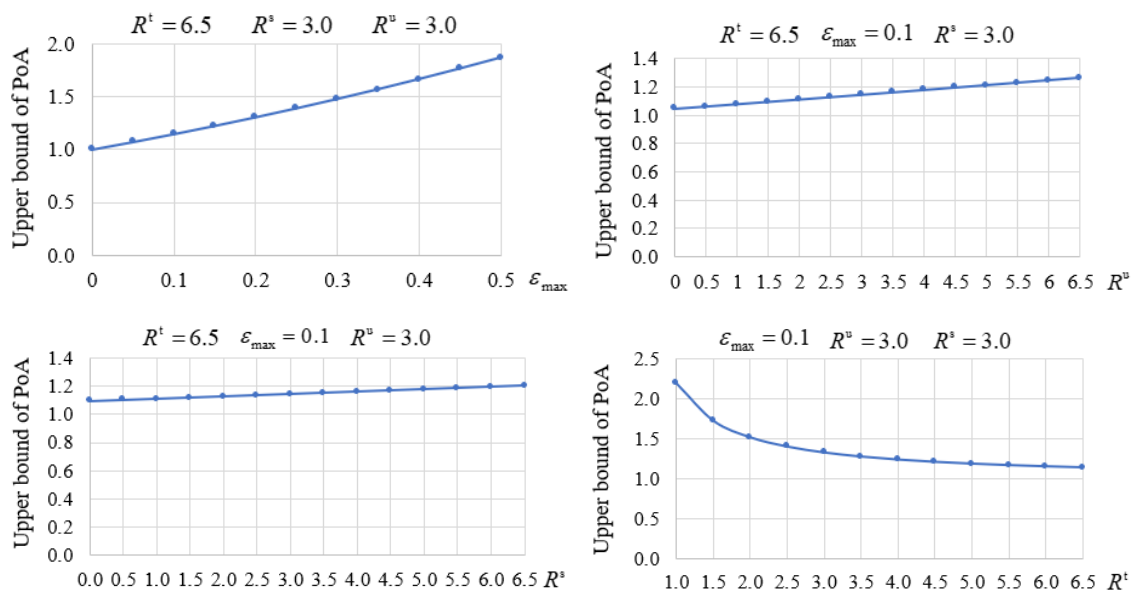


Figure 1. The upper bounds of PoAs given different parameter values

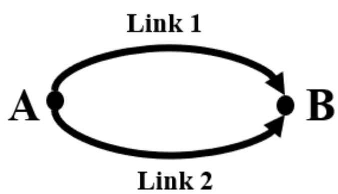
Property 6. The upper bound of the PoA for the capacity expansion RUE-NDP under cost recovery is independent of network topology and travel demands.

Property 7. The upper bound of the PoA for the capacity expansion RUE-NDP under cost recovery is independent of travel time functions.

The proofs of Properties 5, 6, and 7 are straightforward and omitted.

In the following, we present an example to illustrate how to calculate the upper bound of PoA given a design instance.

Example 1:



Design instance:

$R^{\mathrm{t}} = 6.0 \quad R^{\mathrm{u}} = 2.3 \quad R^{\mathrm{s}} = 2.3$

$d_{AB} = 5.0 \quad \kappa_1 = 2.0 \quad \kappa_2 = 3.0$

$t_1 = 2.0 + 0.5x_1/(y_1+2)$

$t_2 = 2.2 + 0.3x_2/(y_2+3)$

$\sigma_1 = 0.2 + 0.01x_1/(y_1+2)$

$\sigma_2 = 0.1 + 0.03x_2/(y_2+3)$

Solution:

$x_1^{RUE} = 2.04 \quad y_1^* = 0.19$

$x_2^{RUE} = 2.96 \quad y_2^* = 0$

$t_1^{RUE} = 2.47 \quad \sigma_1^{RUE} = 0.21$

$t_2^{RUE} = 2.50 \quad \sigma_2^{RUE} = 0.13$

To get an upper bound of the PoA for this design instance, we need the parameters R^{s}, R^{t}, R^{u}, and $\varepsilon_{\max}$. $\varepsilon_1 = 0.21/2.47 = 0.08$ and $\varepsilon_2 = 0.13/2.50 = 0.05$. Take $\varepsilon_{\max} = 0.08$. We also have $R^{\mathrm{t}} = 6.0$, $R^{\mathrm{s}} = 2.3$, and $R^{\mathrm{u}} = 2.3$. The upper bound of PoA is $\left(1+\varepsilon_{\max} R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max} R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2 = 1.10$.

In this example, the upper bound of PoA is independent of the network topology, travel demands, and travel time functions, as indicated in Property 6 and Property 7.

Based on Proposition 1 and Property 5, the following proposition can be directly concluded.

Proposition 2. Denote I as the set of instances in which each instance satisfies the following conditions: 1) the maximum ratio between link travel time standard deviation and mean link travel time does not exceed $\bar{\varepsilon}_{\max}$; 2) the VOR for path travel time does not exceed $\bar{R}^{\mathrm{u}}$; 3) the VOR for system travel time does not exceed $\bar{R}^{\mathrm{s}}$; and 4) the VOT is not less than $\bar{R}^{\mathrm{t}}$. The price of anarchy of I is bounded above:

$$\rho(I) \le \left(1+\bar{\varepsilon}_{\max}\bar{R}^{\mathrm{s}}/\bar{R}^{\mathrm{t}}\right)\left(1+\bar{\varepsilon}_{\max}\bar{R}^{\mathrm{u}}/\bar{R}^{\mathrm{t}}\right)^2. \tag{19}$$

Remark 1. The existence of an upper bound indicates that both $\rho(G,\mathbf{d},\mathbf{t},\theta)$ and $\rho(I)$ for Problem Q are not trivial notions (i.e., the PoA is meaningless if it is unbounded).

Remark 2. The upper bound of the PoA equals one under either of the two conditions: 1) there is no supply uncertainty (i.e., $\overline{\varepsilon}_{\max} = 0$); 2) there are no monetary values in the reliabilities of system travel time and path travel time (i.e., $\overline{R}^{\mathrm{s}} = \overline{R}^{\mathrm{u}} = 0$). The reason is that link marginal mean cost tolls are equivalent to link marginal cost tolls when there is no uncertainty, and charging the link marginal cost tolls drives the travelers to choose paths to minimize the TSTT. In Sub-section 2.2, it is discussed that the PoA for an instance set is intuitively larger than or equal to one. Together with the fact that the upper bound of the PoA is equal to one under either of the conditions, the PoA must equal one. Then, the upper bound of the PoA is equal to the PoA itself.

4.3 Discussions on the upper bound of the PoA

4.3.1 Application of the upper bound

The upper bound of the PoA carries different implications from those of $\rho(G,\mathbf{d},\mathbf{t},\theta)$ and $\rho(I)$. As mentioned in Sub-section 2.2, $\rho(G,\mathbf{d},\mathbf{t},\theta)$ reflects the exact inefficiency of the equilibrium solution given an instance $(G,\mathbf{d},\mathbf{t},\theta)$ and $\rho(I)$ reflects the exact worst-case inefficiency of the equilibrium solutions given a group of instances. The $\rho(G,\mathbf{d},\mathbf{t},\theta)$ and $\rho(I)$ are valuable economic evaluation indexes. The upper bound of the PoA, on the other hand, *is a quick estimate of* $\rho(G,\mathbf{d},\mathbf{t},\theta)$ or $\rho(I)$. Furthermore, computing the upper bound of the PoA only requires the values of a few parameters, which is an advantage when available information is limited. For example, computing $\rho(G,\mathbf{d},\mathbf{t},\theta)$ requires the information of G, $\mathbf{d}$, $\mathbf{t}$, and θ, where θ refers to the information related to link (additional) capacity and link free flow travel time variations. Acquiring this piece of information can be time and resource consuming. On the other hand, computing the upper bound of $\rho(G,\mathbf{d},\mathbf{t},\theta)$ only requires the information of $\varepsilon_{\max}$, R^{t}, R^{u}, and R^{s}, which can be acquired more easily. The system manager or other analysts can quickly and easily estimate the inefficiency of the equilibrium solution and decide if necessary measures are needed to deal with the selfish-routing behavior of travelers.

4.3.2 Comparison with existing studies

We proceed to compare the properties of the upper bound of the PoA for Problem Q to those of the upper bound of the PoA for the RUE-NDP proposed by Szeto and Wang (2015). Property 6 is consistent with the result of Szeto and Wang (2015).

Property 7, however, differs from the result of Szeto and Wang (2015), which indicates that the upper bound of the PoA is dependent on the highest degree of the mean link travel time functions. An implication of Property 7 is that the system manager does not need to acquire the information regarding travel time functions to calculate an *upper bound* of the PoA. Property 7 also implies that the derived upper bound is a bound of the PoAs for design instances in which the travel time functions can take any forms as long as they are differentiable and monotone increasing with respect to the link flows. The upper bound of the PoA proposed by Szeto and Wang (2015), on the other hand, can only bound the PoAs for design instances in which the travel time functions must be polynomial functions.

In the following, we explain why our proposed upper bound of the PoA has Property 7 and the upper bound of the PoA proposed by Szeto and Wang (2015) does not have. Szeto and Wang (2015) assumed specifically that the mean link travel time functions must be polynomial functions with respect to link flows. Our study only assumes that the mean link travel time functions are monotone increasing and differentiable functions with respect to link flows. Szeto and Wang (2015) used the mathematical properties of polynomial mean link travel time functions to derive an upper bound of the inefficiency of the RUE flow pattern, which is dependent on the highest degree of the mean link travel time functions. In our study, because the travelers are charged with link marginal mean cost tolls, we can derive an upper bound of the inefficiency of the worst RUE flow pattern given an optimal capacity expansion vector without knowing the explicit expressions of the mean link travel time functions. Thus, the upper bound of the PoA proposed by Szeto and Wang (2015) is dependent on the link travel time functions whereas ours is not.

5 Conclusion

The study proposed a general definition of the PoA for capacity expansion RUE-NDPs with the following features: 1) the objective function can include total system travel time, travel time reliability, construction cost, environmental cost, and other system performance measures; 2) auxiliary decision variables can be included as long as they do not affect the value of the objective function; 3) the lower level problem can be any type of RUE problems; and 4) additional constraints are incorporated.

This study proposed a capacity expansion RUE-NDP under cost recovery that considers supply uncertainty. The link marginal mean cost tolls are charged upon the travelers, and a cost recovery constraint is incorporated to guarantee that the degree of cost recovery (proposed and defined in this study) is larger than or equal to one. The problem is formulated as a min-max problem

A novel approach to deriving the analytical formula of an upper bound of the PoA is presented. The upper bound is independent of travel time functions, demands, and network topology. The upper bound is related to the travel time variations, the VORs for system travel time and path travel time, and the VOT. The upper bound is a quick estimate of the PoA value when limited information is available.

Acknowledgments

This work was jointly supported by a grant [No. 201711159034] from the University Research Committee of the University of Hong Kong, a grant from the National Natural Science Foundation of China [No. 71771194], and a grant from the Guangzhou Science Technology and Innovation Commission [No. 201707010292]. The authors are grateful to the Editor-in-Chief and two reviewers for their constructive comments.

References

Abdulaal M, LeBlanc LJ (1979) Continuous equilibrium network design models.

Transportation Research Part B 13(1): 19-32

Ben-Ayed O, Boyce DE, Blair III CE (1988) A general bilevel linear programming formulation of the network design problem. Transportation Research Part B 22(4): 311-318

Brownstone D, Small KA (2003) Valuing time and reliability: assessing the evidence from road pricing demonstrations. Transportation Research Part A 39(4): 279-293

Chau CK, Sim KM (2003) The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands. Operations Research Letters 31(5): 327-334

Chen A, Kim J, Zhou Z, Chootinan P (2007) Alpha reliable network design problem. Transportation Research Record 2029: 49-57

Chen A, Zhou Z (2010) The α-reliable mean-excess traffic equilibrium model with stochastic travel times. Transportation Research Part B 44(4): 493-513

Chen A, Zhou Z, Chootinan P, Ryu S, Yang C, Wong SC (2011) Transport network design problem under uncertainty: a review and new developments. Transport Reviews 31(6): 743-768

Chiou S (2005) Bilevel programming for the continuous transport network design problem. Transportation Research Part B 39(4): 361-383

Chootinan P, Wong SC, Chen A (2005) A reliability-based network design problem. Journal of Advanced Transportation 39(3): 247-270

Chow JYJ, Regan AC (2011) Network-based real option models. Transportation Research Part B 45(4): 682-695

Concas S, Kolpakov A (2009) Synthesis of research on value of time and value of reliability. http://www.nctr.usf.edu/pdf/77806.pdf

Correa J, Schulz A, Stier-Moses N (2004) Selfish routing in capacitated networks. Mathematics of Operations Research 29(4): 961-976

Dantzig GB, Maier SF, Harvey RP, Lansdowne ZF, Robinson DW (1979) Formulating and solving the network design problem by decomposition. Transportation Research Part B 13(1): 5-17

Farahani RZ, Miandoabchi E, Szeto WY, Rashidi H (2013) A review of urban transportation network design problems. European Journal of Operational Research 229(2): 281-302

Friesz TL, Anandalingam G, Mehta NJ, Nam K, Shah SJ, Tobin RL (1993) The multiobjective equilibrium network design problem revisited: a simulated annealing approach. European Journal of Operation Research 65(1): 44-57

Guo XL, Yang H, Liu TL (2010) Bounding the inefficiency of logit-based stochastic user equilibrium. European Journal of Operational Research 201(2): 463-469

Han D, Lo HK, Yang H (2008) On the price of anarchy for non-atomic congestion games under asymmetric cost maps and elastic demands. Computers & Mathematics with Applications 56(10): 2737-2743

Han D, Yang H (2008) The multi-class, multi-criterion traffic equilibrium and the efficiency of congestion pricing. Transportation Research Part E 44(5): 753-773

Huang HJ, Liu TL, Guo XL, Yang H (2011) Inefficiency of logit-based stochastic user equilibrium in a traffic network under ATIS. Networks and Spatial Economics 11(2):

255-269
Ji XF, Ban XG, Li MT, Zhang J, Ran B (2017) Non-expected route choice model under risk on stochastic traffic networks. Networks and Spatial Economics 17(3): 777-807
Khani A, Boyles SD (2015) An exact algorithm for the mean-standard deviation shortest path problem. Transportation Research Part B 81: 252-266
Koutsoupias E, Papadimitriou C (1999) Worst-case equilibria. In: Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS): Trier, Germany, Lecture Notes in Computer Science, vol. 1563, Springer, Berlin, 404-413
LeBlanc LJ, Boyce DE (1986) A bilevel programming algorithm for exact solution of the network design problem with user-optimal flows. Transportation Research Part B 20(3): 259-265
Liu ZY, Yi W, Wang SA, Chen J (2017) On the uniqueness of user equilibrium flow with speed limit. Networks and Spatial Economics 17(3): 763-775
Lo HK, Luo XW. Siu BWY (2006) Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. Transportation Research Part B 40(9): 792-806
Lo HK, Szeto WY (2009) Time-dependent transport network design under cost-recovery. Transportation Research Part B 43(1): 142-158
Meng Q, Yang H (2002) Benefit distribution and equity in road network design. Transportation Research Part B 36(1): 19-35
Ng MW, Waller ST (2009) Reliable system-optimal network design: convex mean-variance model with implicit chance constraints. Transportation Research Record 2090: 68-74
Roughgarden T (2005) Selfish routing and price of anarchy. MIT Press. Cambridge, MA
Roughgarden T, Tardos E (2002) How bad is selfish routing? Journal of the ACM 49(2): 236-259
Shao H, Lam WHK, Tam ML (2006) A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. Networks and Spatial Economics 6(3): 173-204
Small KA, Yan J (2001) The value of “value pricing” of roads: second-best pricing and product differentiation. Journal of Urban Economics 49: 310-336
Sumalee A, Luathep P, Lam WHK, Connors RD (2009) Transport network capacity evaluation and design under demand uncertainty. Transportation Research Record 2090: 17-28
Szeto WY, Jaber XQ, O’Mahony M (2010) Time-dependent discrete network design frameworks considering land use. Computer-Aided Civil and Infrastructure Engineering 25(6): 411-426
Szeto WY, Lo HK (2005) Strategies for road network design over time: robustness under uncertainty. Transportmetrica 1(1): 47-63
Szeto WY, Lo HK (2008) Time-dependent transport network improvement and tolling strategies. Transportation Research Part A 42(2): 376-391
Szeto WY, Wang B (2015) Price of anarchy for reliability-based traffic assignment and network design. Transportmetrica A: Transport Science 11(7): 603-635
Szeto WY, Wang B (2016) Reliable network design under supply uncertainty with probabilistic guarantees. Transportmetrica A: Transport Science 12(6): 504-532

Szeto WY, Wang Y, Wong SC (2014) The chemical reaction optimization approach to solving the environmentally sustainable network design problem. Computer-Aided Civil and Infrastructure Engineering 29(2): 140-158

Tilahun N, Levinson DM (2009) Value of time comparisons in the presence of unexpected delay. Travel Demand Management and Road User Pricing: Success, Failure and Feasibility, Wafaa Saleh & Gerd Sammer, eds., 173-184, Ashgate Publishers.

Wang B, Szeto WY (2018) Reliability-based user equilibrium in a transport network under the effects of speed limits and supply uncertainty. Applied Mathematical Modelling 56: 186-201

Wang CL, Doan XV, Chen B (2014) Price of anarchy for non-atomic congestion games with stochastic demands. Transportation Research Part B 70: 90-111

Wardrop JG (1952) Some theoretical aspects of road traffic research. ICE Proceedings: Engineering Divisions 1(3): 325-362

Wu X (2015) Study on mean-standard deviation shortest path problem in stochastic and time-dependent networks: A stochastic dominance based approach. Transportation Research Part B 80(9): 275-290

Wu X, Nie Y (2011) Modeling heterogeneous risk-taking behavior in route choice: a stochastic dominance approach. Transportation Research Part A 45(9): 896-915

Xiao F, Yang H, Han D (2007) Competition and efficiency of private toll roads. Transportation Research Part B 41(3): 292-308

Yang H (1997) Sensitivity analysis for the elastic-demand network equilibrium problem with applications. Transportation Research Part B 31(1): 55-70

Yang H, Bell MGH (1998) Models and algorithms for road network design: a review and some new developments. Transport Reviews 18(3): 257-278

Yang H, Meng Q (2002) A note on “highway pricing and capacity choice in a road network under a build-operate-transfer scheme”. Transportation Research Part A 36(7): 659-663

Yin YF, Madanat SM, Lu XY (2009) Robust improvement schemes for road networks under demand uncertainty. European Journal of Operational Research 198(2): 470-479

Zhao XM, Wan CH, Bi J (2018) Day-to-day assignment models and traffic dynamics under information provision. Networks and Spatial Economics, https://doi.org/10.1007/s11067-018-9386-1

Appendix A. Lemma 1 and its proof

Lemma 1. For any link flow pattern $\mathbf{v}' = (v'_a)_{a\in A}$, the following inequality holds:

$$\begin{aligned}&\sum_{a\in A}\Big(R^{\mathrm{t}} t_a(v_a^{RUE}, y_a^{RUE}) + \tau_a(v_a^{RUE}, y_a^{RUE}) - R^{\mathrm{t}} t_a(v'_a, y_a^{RUE})\Big)v'_a \\ &\le \sum_{a\in A}\tau_a(v_a^{RUE}, y_a^{RUE})\cdot v_a^{RUE},\end{aligned} \quad \text{(A.1)}$$

where v_a^{RUE} and y_a^{RUE} denote the entries of $\mathbf{v}^{RUE}(\mathbf{f}^{RUE})$ and $\mathbf{y}^{RUE}$, respectively.

Proof. For an individual link $a \in A$, consider the following maximization problem:

$$\min_{x_a \geq 0} \bar{Z}_a(x_a) = R^{\mathrm{t}}\left(t_a(v_a^{RUE}, y_a^{RUE}) + v_a^{RUE} dt_a(v_a^{RUE}, y_a^{RUE})/dv_a - t_a(x_a, y_a^{RUE})\right) x_a .$$

The first order derivative of $\bar{Z}_a(x_a)$ with respect to x_a is

$$d\bar{Z}_a(x_a)/dx_a =$$
$$R^{\mathrm{t}}\left(t_a(v_a^{RUE}, y_a^{RUE}) + v_a^{RUE} dt_a(v_a^{RUE}, y_a^{RUE})/dv_a - t_a(x_a, y_a^{RUE}) - x_a dt_a(x_a, y_a^{RUE})/dv_a\right).$$

Because of the properties of the link travel time function and the marginal link cost toll function, the following hold: $d\bar{Z}_a(x_a)/dx_a > 0$ for $0 \leq x_a < v_a^{RUE}$; $d\bar{Z}_a(x_a)/dx_a = 0$ for $x_a = v_a^{RUE}$, and $d\bar{Z}_a(x_a)/dx_a < 0$ for $x_a > v_a^{RUE}$. The objective function $\bar{Z}_a(x_a)$ is strictly increasing on $\left[0, v_a^{RUE}\right]$ and strictly decreasing on $\left(v_a^{RUE}, +\infty\right)$. If v_a^{RUE} equals zero, $d\bar{Z}_a(x_a)/dx_a = 0$ at $x_a = 0$ and $d\bar{Z}_a(x_a)/dx_a < 0$ for $x_a > 0$. The function $\bar{Z}_a(x_a)$ is strictly decreasing on $\left[0, +\infty\right)$. The global maximum point x_a^* of the objective function exists and is unique, and satisfies the condition: $d\bar{Z}_a(x_a^*)/dx_a = 0$, i.e., $x_a^* = v_a^{RUE}$.

Substituting the global maximum point x_a^* into the objective function $\bar{Z}_a(x_a)$, the maxima of the objective function is

$$\begin{aligned}\bar{Z}_a(x_a^*) &= R^{\mathrm{t}}\left(t_a(v_a^{RUE}, y_a^{RUE}) + v_a^{RUE} dt_a(v_a^{RUE}, y_a^{RUE})/dv_a - t_a(v_a^{RUE}, y_a^{RUE})\right) v_a^{RUE} \\ &= R^{\mathrm{t}} v_a^{RUE} v_a^{RUE} \cdot dt_a(v_a^{RUE}, y_a^{RUE})/dv_a = \tau_a(v_a^{RUE}, y_a^{RUE}) \cdot v_a^{RUE}.\end{aligned}$$

Thus, given a feasible link flow v'_a, the following inequality holds:

$$\begin{aligned}\bar{Z}_a(v'_a) &= R^{\mathrm{t}}\left(t_a(v_a^{RUE}, y_a^{RUE}) + v_a^{RUE} dt_a(v_a^{RUE}, y_a^{RUE})/dv_a - t_a(v'_a, y_a^{RUE})\right) v'_a \\ &= \left(R^{\mathrm{t}} t_a(v_a^{RUE}, y_a^{RUE}) + \tau_a(v_a^{RUE}, y_a^{RUE}) - R^{\mathrm{t}} t_a(v'_a, y_a^{RUE})\right) v'_a \qquad \text{(A.2)} \\ &\leq \tau_a(v_a^{RUE}, y_a^{RUE}) \cdot v_a^{RUE}.\end{aligned}$$

Condition (A.2) holds for any individual link in the road network. Summing up condition (A.2) over all links on a path, the result (A.1) in the lemma is obtained. ■

Appendix B. Upper bounds of TSTCB and sum of individual path travel cost budgets

Based on the formula relating the path and link travel time standard deviation, the path travel time standard deviation is smaller than or equal to the sum of link travel time standard deviations of links on that path, i.e., $\varsigma_p \leq \sum_{a \in A} \sigma_a \delta_p^a$. Similarly, $\sigma\left[\widehat{TSTT}\right] \leq \sum_{a \in A} \sigma_a v_a$. According to the definition of $\varepsilon_{\max}$, we have $\varsigma_p \leq \sum_{a \in A} \varepsilon_{\max} t_a \delta_p^a$ and $\sigma\left[\widehat{TSTT}\right] \leq \sum_{a \in A} \varepsilon_{\max} t_a v_a$.

Because $\sigma\left[\widehat{TSTT}\right] \le \sum_{a\in A} \varepsilon_{\max} t_a v_a$, it can easily be proved that the TSTCB has an upper bound, which is the mean TSTT multiplied by a number:

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}(\mathbf{f}),\mathbf{y}) \le \left(R^{\mathrm{t}} + \varepsilon_{\max} R^{\mathrm{s}}\right) \sum_{a\in A} t_a(v_a, y_a)\cdot v_a . \tag{B.1}$$

Similar to the sum of individual path travel cost budgets, we have

$$\sum_{p\in P} f_p b_p(\mathbf{v}(\mathbf{f}),\mathbf{y}) \le \left(R^{\mathrm{t}} + \varepsilon_{\max} R^{\mathrm{u}}\right) \sum_{a\in A} t_a(v_a, y_a)\cdot v_a . \tag{B.2}$$

Note that $\sum_{p\in P} f_p b_p(\mathbf{v}(\mathbf{f}),\mathbf{y}) = R^{\mathrm{t}} \sum_{a\in A} t_a(v_a, y_a)\cdot v_a + R^{\mathrm{u}} \sum_{p\in P} f_p \cdot \varsigma_p$.

Appendix C. Proof of Property 1

Proof. Assume $\mathbf{v}'''(\mathbf{f}''') = (v_a''')_{a\in A}$ is the link flow pattern of the path flow pattern $\mathbf{f}''' = (f_p''')_{p\in P}$ that minimizes the sum of individual path travel cost budgets. Let ς_p^{RUE} and ς_p''' be the path travel time standard deviations of $\mathbf{f}^{RUE}$ and $\mathbf{f}'''$, respectively. Let $\tau_a^{RUE} = \tau_a(v_a^{RUE}, y_a^{RUE})$, $\tau_a''' = \tau_a(v_a''', y_a^{RUE})$, $t_a^{RUE} = t_a(v_a^{RUE}, y_a^{RUE})$, $t_a''' = t_a(v_a''', y_a^{RUE})$, $q_p^{RUE} = \sum_{a\in A} t_a^{RUE}\delta_p^a$, $q_p''' = \sum_{a\in A} t_a'''\delta_p^a$, $\tilde{b}_p^{RUE} = b_p^{RUE} + \sum_{a\in A} \tau_a^{RUE}\delta_p^a$, and $\tilde{b}_p''' = b_p''' + \sum_{a\in A} \tau_a'''\delta_p^a$.

Because $\mathbf{f}^{RUE}$ is the RUE path flow pattern, the following inequality holds: $\sum_{p\in P}\left(f_p''' - f_p^{RUE}\right)\tilde{b}_p^{RUE} \ge 0$ (for details, see the solution method in Sub-section 2.7 in the study of Szeto and Wang 2016), which is equivalent to $\sum_{p\in P} \tilde{b}_p^{RUE} f_p^{RUE} \le \sum_{p\in P} \tilde{b}_p^{RUE} f_p'''$. Subtracting $\sum_{p\in P} \tilde{b}_p''' f_p'''$ from both sides of the above inequality, we obtain $\sum_{p\in P} \tilde{b}_p^{RUE} f_p^{RUE} - \sum_{p\in P} \tilde{b}_p''' f_p''' \le \sum_{p\in P}\left(\tilde{b}_p^{RUE} - \tilde{b}_p'''\right) f_p'''$, which can be rewritten as

$$\begin{aligned}&\sum_{p\in P} b_p^{RUE} f_p^{RUE} + \sum_{p\in P}\left(\sum\nolimits_{a\in A} \tau_a^{RUE}\delta_p^a\right) f_p^{RUE} - \sum_{p\in P} b_p''' f_p''' - \sum_{p\in P}\left(\sum\nolimits_{a\in A} \tau_a^{RUE}\delta_p^a\right) f_p''' \le \\ &R^{\mathrm{t}} \sum_{p\in P}\left(q_p^{RUE} - q_p'''\right) f_p''' + R^{\mathrm{u}} \sum_{p\in P}\left(\varsigma_p^{RUE} - \varsigma_p'''\right) f_p''' + \sum_{p\in P}\left(\sum\nolimits_{a\in A} \tau_a^{RUE}\delta_p^a - \sum\nolimits_{a\in A} \tau_a'''\ \delta_p^a\right) f_p''',\end{aligned}$$

or

$$\begin{aligned}&\sum_{p\in P} b_p^{RUE} f_p^{RUE} - \sum_{p\in P} b_p''' f_p''' \\ &\le \left[R^{\mathrm{t}} \sum_{p\in P}\left(q_p^{RUE} - q_p'''\right) f_p''' + \sum_{p\in P}\left(\sum\nolimits_{a\in A} \tau_a^{RUE}\delta_p^a\right) f_p''' - \sum_{p\in P}\left(\sum\nolimits_{a\in A} \tau_a^{RUE}\delta_p^a\right) f_p^{RUE}\right] \\ &+ R^{\mathrm{u}} \sum_{p\in P}\left(\varsigma_p^{RUE} - \varsigma_p'''\right) f_p'''.\end{aligned} \tag{C.1}$$

For the term in the square bracket on the right side of (C.1), we have: $R^{\mathrm{t}}\sum_{p\in P}\left(q_p^{RUE}-q_p'''\right)f_p'''=R^{\mathrm{t}}\sum_{a\in A}\left(t_a^{RUE}-t_a'''\right)v_a'''$, $\sum_{p\in P}\left(\sum_{a\in A}\tau_a^{RUE}\delta_p^a\right)f_p'''=\sum_{a\in A}\tau_a^{RUE}v_a'''$, and $\sum_{p\in P}\left(\sum_{a\in A}\tau_a^{RUE}\delta_p^a\right)f_p^{RUE}=\sum_{a\in A}\tau_a^{RUE}v_a^{RUE}$. Thus, the term in the square bracket in (C.1) can be expressed in terms of link-based variables:

$$\begin{aligned}&\left[R^{\mathrm{t}}\sum_{p\in P}\left(q_p^{RUE}-q_p'''\right)f_p'''+\sum_{p\in P}\left(\sum_{a\in A}\tau_a^{RUE}\delta_p^a\right)f_p'''-\sum_{p\in P}\left(\sum_{a\in A}\tau_a^{RUE}\delta_p^a\right)f_p^{RUE}\right]=\\&\sum_{a\in A}\left(R^{\mathrm{t}}t_a^{RUE}-R^{\mathrm{t}}t_a'''+\tau_a^{RUE}\right)v_a'''-\sum_{a\in A}\tau_a^{RUE}v_a^{RUE}.\end{aligned}\tag{C.2}$$

According to Lemma 1 in Appendix A, the first term on the right side of inequality (C.2) is smaller than or equal to $\sum_{a\in A}\tau_a^{RUE}v_a^{RUE}$. Thus, the right side of (C.2) is smaller than or equal to zero. Because the term in the square bracket in (C.1) is smaller than or equal to zero, the left side of (C.1) is smaller than or equal to the second term on the right side of (C.1):

$$\sum_{p\in P}b_p^{RUE}f_p^{RUE}-\sum_{p\in P}b_p'''f_p'''\le 0+R^{\mathrm{u}}\sum_{p\in P}\left(\varsigma_p^{RUE}-\varsigma_p'''\right)f_p'''.\tag{C.3}$$

It is assumed that the mapping $\boldsymbol{\varsigma}=\left(\varsigma_p\right)_{\forall p\in P}$ is monotone in terms of path flow $\mathbf{f}$. Thus, the following holds:

$$R^{\mathrm{u}}\sum_{p\in P}\left(\varsigma_p'''-\varsigma_p^{RUE}\right)\left(f_p^{RUE}-f_p'''\right)\le 0\text{, or equivalently,}$$

$$R^{\mathrm{u}}\sum_{p\in P}\varsigma_p'''f_p^{RUE}+R^{\mathrm{u}}\sum_{p\in P}\left(\varsigma_p^{RUE}-\varsigma_p'''\right)f_p'''\le R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{RUE}f_p^{RUE}.$$

Eliminating the non-negative term $R^{\mathrm{u}}\sum_{p\in P}\varsigma_p'''f_p^{RUE}$ from the above inequality, the inequality still holds, i.e.,

$$R^{\mathrm{u}}\sum_{p\in P}\left(\varsigma_p^{RUE}-\varsigma_p'''\right)f_p'''\le R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{RUE}f_p^{RUE}.\tag{C.4}$$

Based on inequalities (C.3) and (C.4), the following is true:

$$\sum_{p\in P}b_p^{RUE}f_p^{RUE}-\sum_{p\in P}b'''f_p'''\le R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{RUE}f_p^{RUE}.\tag{C.5}$$

Based on (B.2), the following inequality holds:

$$\sum_{p\in P}f_pb_p\le\left(R^{\mathrm{t}}+\varepsilon_{\max}R^{\mathrm{u}}\right)\cdot\frac{1}{R^{\mathrm{t}}}\cdot\left(\sum_{p\in P}f_pb_p-R^{\mathrm{u}}\sum_{p\in P}f_p\cdot\varsigma_p\right)\text{, or equivalently,}$$

$$R^{\mathrm{u}}\sum_{p\in P}f_p\cdot\varsigma_p\le\left(1-\frac{R^{\mathrm{t}}}{R^{\mathrm{t}}+\varepsilon_{\max}R^{\mathrm{u}}}\right)\sum_{p\in P}f_pb_p.$$

Based on the above, the following inequality holds:

$$R^{\mathrm{u}}\sum_{p\in P}\varsigma_p^{RUE} f_p^{RUE} \le \left(R^{\mathrm{u}}\varepsilon_{\max}\Big/\left(R^{\mathrm{t}}+R^{\mathrm{u}}\varepsilon_{\max}\right)\right)\sum_{p\in P} b_p^{RUE} f_p^{RUE} .$$

The right side of the above inequality is an upper bound of the right side of (C.5) and thus is an upper bound of the left side of (C.5), which gives

$$\sum_{p\in P} b_p^{RUE} f_p^{RUE} - \sum_{p\in P} b_p''' f_p''' \le \left(R^{\mathrm{u}}\varepsilon_{\max}\Big/\left(R^{\mathrm{t}}+R^{\mathrm{u}}\varepsilon_{\max}\right)\right)\sum_{p\in P} b_p^{RUE} f_p^{RUE} ,$$

which further gives

$$\sum_{p\in P} b_p^{RUE} f_p^{RUE} \Big/ \sum_{p\in P} b_p''' f_p''' \le 1\Big/\left(1-\left(R^{\mathrm{u}}\varepsilon_{\max}\Big/\left(R^{\mathrm{t}}+R^{\mathrm{u}}\varepsilon_{\max}\right)\right)\right) = 1+\left(R^{\mathrm{u}}\varepsilon_{\max}\Big/R^{\mathrm{t}}\right).$$

This completes the proof. ■

Appendix D. **Proof of Property 2**

Proof. Assume $\mathbf{v}''(\mathbf{f}'') = \left(v_a''\right)_{a\in A}$ is the link flow pattern of the path flow pattern $\mathbf{f}'' = \left(f_p''\right)_{p\in P}$ that minimizes the TSTCB given $\mathbf{y}^{RUE}$. Let $b_p'' = b_p(\mathbf{v}''(\mathbf{f}''), \mathbf{y}^{RUE})$.

By definition, the sum of individual path travel cost budgets is larger than or equal to the monetary value of mean TSTT, i.e.,

$$R^{\mathrm{t}}\sum_{a\in A} t_a^{RUE} v_a^{RUE} \le \sum_{p\in P} b_p^{RUE} f_p^{RUE} .$$

Multiplying both sides of the above inequality by $\left(1+\varepsilon_{\max}R^{\mathrm{s}}\big/R^{\mathrm{t}}\right)$, the inequality still holds. That is,

$$\left(R^{\mathrm{t}}+\varepsilon_{\max}R^{\mathrm{s}}\right)\sum_{a\in A} t_a^{RUE} v_a^{RUE} \le \left(1+\varepsilon_{\max}R^{\mathrm{s}}\big/R^{\mathrm{t}}\right)\sum_{p\in P} b_p^{RUE} f_p^{RUE} .$$

The left side of above inequality is an upper bound of the TSTCB according to (B.1) in Appendix B. Thus, the right side of the above inequality is larger than or equal to $TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}), \mathbf{y}^{RUE})$, i.e.,

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}), \mathbf{y}^{RUE}) \le \left(1+\varepsilon_{\max}R^{\mathrm{s}}\big/R^{\mathrm{t}}\right)\sum_{p\in P} b_p^{RUE} f_p^{RUE} . \tag{D.1}$$

Similarly, the following inequality holds:

$$\left(R^{\mathrm{t}}+\varepsilon_{\max}R^{\mathrm{u}}\right)\sum_{a\in A} t_a'' v_a'' \le \left(1+\varepsilon_{\max}R^{\mathrm{u}}\big/R^{\mathrm{t}}\right) TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}''(\mathbf{f}''), \mathbf{y}^{RUE}).$$

The left side of the above inequality is an upper bound of the sum of individual path travel cost budgets according to (B.2) in Appendix B. Thus, the right side of the above inequality is larger than or equal to $\sum_{p\in P} b_p'' f_p''$, which further gives

$$TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}''(\mathbf{f}''), \mathbf{y}^{RUE}) \ge \left(\sum_{p\in P} b_p'' f_p''\right)\Big/\left(1+\varepsilon_{\max}R^{\mathrm{u}}\big/R^{\mathrm{t}}\right). \tag{D.2}$$

Dividing the left side of (D.1) by the left side of (D.2), and dividing the right side of (D.1) by the right side of (D.2), we obtain the following inequality:

$$\frac{TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}^{RUE}(\mathbf{f}^{RUE}),\mathbf{y}^{RUE})}{TSTCB_{R^{\mathrm{t}},R^{\mathrm{s}}}(\mathbf{v}''(\mathbf{f}''),\mathbf{y}^{RUE})} \le \frac{\sum_{p\in P} b_p^{RUE} f_p^{RUE}}{\sum_{p\in P} b_p'' f_p''}\left(1+\varepsilon_{\max} R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max} R^{\mathrm{u}}/R^{\mathrm{t}}\right). \quad \text{(D.3)}$$

In (D.3), $\sum_{p\in P} b_p'' f_p''$ is larger than $\sum_{p\in P} b_p''' f_p'''$ defined in Property 1, because $\sum_{p\in P} b_p''' f_p'''$ is the minimum sum of individual path travel cost budgets given $\mathbf{y}^{RUE}$. Thus,

$$\sum_{p\in P} b_p^{RUE} f_p^{RUE} \Big/ \sum_{p\in P} b_p'' f_p'' \le \sum_{p\in P} b_p^{RUE} f_p^{RUE} \Big/ \sum_{p\in P} b_p''' f_p''' .$$

Together with Property 1, we obtain the following inequality:

$$\sum_{p\in P} b_p^{RUE} f_p^{RUE} \Big/ \sum_{p\in P} b_p'' f_p'' \le 1+\varepsilon_{\max} R^{\mathrm{u}}/R^{\mathrm{t}} . \quad \text{(D.4)}$$

Inequalities (D.3) and (D.4) indicate that the left side of (D.3) is smaller than or equal to $\left(1+\varepsilon_{\max} R^{\mathrm{s}}/R^{\mathrm{t}}\right)\left(1+\varepsilon_{\max} R^{\mathrm{u}}/R^{\mathrm{t}}\right)^2$. This completes the proof. ■