# Combined Decision Making of Congestion Pricing and Capacity Expansion: Genetic Algorithm Approach

Wei (David) Fan[1] and Zegeye Gurmu[2]

**Abstract:** This paper presents a solution methodology that can be used to determine the optimal solution for the combined congestion pricing and capacity expansion problems. A bilevel genetic algorithm (GA)-based optimization solution methodology is proposed to determine the optimal toll location, toll rate, percentage capacity expansion, and location for the expansion simultaneously. The upper-level subprogram minimizes the total system travel cost given certain budget and toll constraints. The lower-level subprogram is a user equilibrium problem where all users try to find the route that minimizes their own travel cost (or time). The budget constraint is handled using a penalty parameter. The demand is assumed to be fixed and given a priori. The proposed GA model is applied to Sioux Falls network, which has 76 links and 24 origin-destination (OD)-pairs, assuming homogeneous users. The optimal solution is thus identified. Sensitivity analyses are conducted for the budget and penalty parameter. The proposed methodology will be a very useful tool for transportation network planners for allocation of budgets and prioritization of links for improvements and congestion pricing. **DOI: 10.1061/(ASCE)TE.1943-5436.0000695.** 

**Author keywords:** Traffic congestion; Traffic management; Pricing; Tolls; Capacity expansion.



## Introduction

Traffic congestion has been continually growing and posing a threat to the quality of life of people in many countries over the past few decades. Congestion in general results in a decrease in accessibility and mobility, travel time loss, and air pollution. So far, different solutions have been suggested, including demand side (e.g., congestion pricing, traffic management) and supply side (e.g., constructing more roads, capacity expansion), or their integration, for mitigation of congestion. The problem of road congestion has long been a research subject, and there are well-established mathematical models for transportation networks using cost and demand functions and the behavior of road users. Among the aforementioned mitigation examples, congestion pricing from the demand side and capacity expansion from the supply side have been studied by many authors (Ordonez and Zhao 2007; Mathew and Sharma 2009; Yang and Zhang 2003; Hearn and Ramana 1998; Bar-Gera et al. 2013). In general, capacity expansion and congestion pricing problems are separately considered and researched as transportation network design problems (NDP) in much literature (Lo and Szeto 2003; Miandoabchi and Farahani 2011). Readers are referred to Farahani et al. (2013) for a comprehensive review of several variations of network design problems.

Capacity expansion has been the common answer to demand growth in the last decades. Network capacity expansion problems involve determining the selection of location of links and how much additional capacity is to be expanded to each of these existing links to minimize the total system costs under a limited budget, while accounting for the route choice behavior of network users (Zhang and Gao 2009). Different approaches have been developed so far for solving such kinds of problems. For example, Mathew and Sharma (2009) investigated a solution to the continuous network design problem using the application of a genetic algorithm to find optimal capacity expansion for a large city network. Miandoabchi and Farahani (2011) addressed the problem of designing of street directions and lane additions in urban road networks using the concept of reserve capacity. The proposed problems were modeled as mixed-integer, bilevel mathematical problems. Li et al. (2012) presented a global optimization method for continuous network design problems. More recently, Miandoabchi et al. (2013) tackled the problem of designing urban road networks in a multiobjective decision-making framework to find the optimal combination of one-way and two-way links, the optimal selection of network capacity expansion projects, the optimal lane allocations on two-way links to optimize the reserve capacity of the network, and two new travel-time-related performance measures. In their study also, the proposed variations were formulated as mixed-integer programming problems with equilibrium constraints. However, most of these studies attempted to address the optimal capacity expansion problem on predetermined links. Very recently, Wang et al. (2013) presented two global optimization algorithms, including the system optimal (SO)-relaxation and user equilibrium (UE)-reduction to address the discrete network design problem.

Likewise, congestion pricing has also long been recognized as a potential way of reducing traffic congestion and studied by many researchers (Yang and Zhang 2003; Hearn and Ramana 1998; Yang and Huang 2005; Verhoef et al. 1996). The so-called second-best pricing scheme, where only a subset of links is subjected to toll charges, has lately received much attention (Yang and Huang 2005; Verhoef et al. 1996). The congestion pricing problem involves the determination of toll levels and/or toll locations for a given network. However, not much attention has been given to the combined determination of toll levels and toll locations. Several methods

[1]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of North Carolina at Charlotte, EPIC Building, Room 3261, 9201 University City Blvd., Charlotte, NC 28223 (corresponding author). E-mail: wfan7@uncc.edu

[2]Traffic Engineer, Teague Nall & Perkins, Inc., 1100 Macon St., Fort Worth, TX 76102. E-mail: zgurmu@tnpinc.com

Note. This manuscript was submitted on November 3, 2013; approved on March 25, 2014; published online on April 29, 2014. Discussion period open until September 29, 2014; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering***, © ASCE, ISSN 0733-947X/04014031(9)/$25.00.

have been proposed to deal with the congestion pricing problem in the past few decades. For example, Verhoef (2002) proposed a derivative-based approach to solve an optimal toll level and location problem. Cree et al. (1998) developed the genetic algorithm (GA)-based method to solve only the optimal toll problem. Shepherd and Sumalee (2004) developed an alternative GA-based approach for finding optimal toll levels for a predefined set of chargeable links and for finding optimal toll locations. A brief review on different methods to solving optimal toll problems can be found elsewhere (Shepherd and Sumalee 2004). Ekström et al. (2012) developed a mixed-integer linear approximation approach to addressing the toll design problem of finding the toll locations and levels in a congestion pricing scheme. Zhang and Sun (2013) formulated the cordon pricing design problem with elastic demand as a mathematical program with complementarity constraints (MPCC) to simultaneously optimize the cordon locations and cordon-specific toll levels to maximize total social welfare.

As was previously noted, considerable research efforts have been directed towards congestion pricing and capacity expansion problems for which each problem is treated separately. It is well known that there exists an imperative relationship between these two problems. However, much transportation network design literature fails to highlight the crucial interaction between road pricing levels and road capacity investment. Although optimal road prices are primarily meant to restrict traffic volumes within existing road capacity, the need for prices to balance supply of capacity and travel demand should not be overlooked. As such, the main objective of this research is to present a solution methodology that is used to determine the optimal combination of continuous capacity expansion and congestion charges along with their corresponding locations, simultaneously.

It is not normally easy to solve such network design problem with traditional network optimization methods because there are too many combinations for choosing the required number of toll or capacity links, even from a limited number of subsets. Determining toll and capacity addition levels on these links is an even more complex task because of their continuous variables characteristics. Therefore, there is a need to find a method that can go through as many combinations as possible to search for the optimal combination. Genetic algorithm is capable of doing this, because it has a favorable procedure of natural selection. Genetic algorithm-based approached have been applied to address toll design and capacity expansion problems by a few researchers (Mathew and Sharma 2009; Yang and Zhang 2003). The network design problem is traditionally formulated as a bilevel optimization problem (Yang and Bell 1998). In most studies, the upper-level problem is to minimize the total system's travel time, whereas the lower level problem minimizes the individual drivers' travel time by the equilibrium traffic assignment problem. Therefore, in this research, an attempt is also made to explore a bilevel GA-based optimization model to solve the second-best optimal pricing and continuous capacity investment problem simultaneously. In this paper, for simplicity, the travel demand is assumed to be fixed and given a priori.

The rest of this paper is organized as follows: "Model Formulation" presents model formulation of the bilevel programming; "Solution Methodology" proposes the solution methodology for the combined pricing and capacity expansion problem; and "Numerical Experimentation" gives network experiments and discusses numerical results. Finally in "Summary and Future Research," a summary and discussion of future research directions conclude this paper.

## Model Formulation

### *Mathematical Notation*

Network design problem can be described in terms of nodes, links, and routes. Consider a connected network with a directed graph $G = \{N, K\}$ consisting of a finite set of $N$ nodes and $K$ links (arcs), such that link $k \in K$, which connect pairs of nodes. In order to formulate the model, the notations described in the Notation section are used.

### *Bilevel Model Formulation*

Basically, bilevel programming formulation involves two players at different levels: the leader and the follower. The two levels have their own decision variables and objectives, and they make an attempt to optimize their own objectives in sequence. The general bilevel programming formulation can be formulated as follows:

$$(\text{UP}) \min_{x \in X} F[x, z]$$

$$\text{Subject to } H[x, z] \leq 0$$

$$(\text{LP}) \min_{z \in Z} f[x, z]$$

$$\text{Subject to } h[x, z] \leq 0$$

where $x$ and $z$ are called leader's and follower's vectors, whereas $F$ and $f$ are called leader's and follower's objective functions, respectively. $H$ and $h$ = constrained set of the upper- and lower-level programs, respectively. In bilevel programming, the leader moves first by choosing a vector $x$ to optimize $F$. For each fixed $x$, the lower level optimizes its objective function $f$ by selecting a vector $z$, which is an optimal solution to the upper-level programming.

A bilevel programming model for determining optimal pricing and capacity investment is proposed in this paper. A bilevel formulation is important to design effective and efficient algorithms to solve network design problems such as congestion pricing and capacity expansion problems. The upper-level program is to minimize the total system travel cost, and the lower-level program is the traffic user equilibrium model in terms of generalized travel cost. The bilevel formulation has been adopted by different authors in the past year to solve different discrete network design problems (Yang 1996; Yang and Bell 1997; Zhang and Yang 2004; Clegg et al. 2001). In the next section, the proposed bilevel optimization model is presented.

#### **Proposed Bilevel Optimization Model**

For simplicity, in the propose optimization model, the value of time for all users in the transportation network is assumed to be the same. The upper-level program that minimizes the total system travel cost is expressed as follows:

$$\min\left\{\sum_{k\in K} t_k[v_k(y,\mu)]v_k(y,\mu)\right\} \tag{1}$$

Subject to

$$y_k^{\min} \leq y_k \leq y_k^{\max} \qquad \sum_{k\in \bar{K}} g_k(\mu_k) \leq B \qquad \mu_k^{\min} \leq \mu_k \leq \mu_k^{\max}$$

where $\sum_{k\in \bar{K}} g_k(\mu_k)$ = total capacity expansion expenditure; and $v_k(y,\mu)$ = the solution for the following lower-level program, which represents the deterministic user equilibrium with given fixed O-D demand:

$$\min\left(\sum_{k\in K}\int_0^{v_k} C_k(x, y_k, \mu_k)dx\right) \tag{2}$$

Subject to

$$v_k = \sum_{w\in W}\sum_{p\in P_w} f_p^w * \delta_{kp}^w, \quad k\in K$$

$$\sum_{p\in P_w} f_p^w = q_w, \quad w\in W$$

$$f_p^w \geq 0, \quad p\in P_w, \quad w\in W$$

where the expression $C_k(x, y_k, \mu_k)$ in the objective function of the lower-level program stands for the generalized link cost function, which is usually expressed by the following most widely used function called the Bureau of Public Roads (BP) function as

$$t_{0k}\left[1+\alpha\left(\frac{V_k}{\text{Cap}_k(1+\mu_k)}\right)^{\beta}\right]+\frac{y_k}{\text{VOT}} \tag{3}$$

where $t_{ok}$ = free flow travel time on link $k\in K$; $\text{Cap}_k$ = initial capacity of link $k\in K$; and $\alpha$, $\beta$ = empirically determined coefficients.

Common values for the coefficients are $\alpha = 0.15$ and $\beta = 4$. It is normally assumed that capacity at this value of $\alpha$ in the preceding formula represents the level of traffic intensity whereby the travel time on the link is 15% higher than the travel time at free flow.

To handle the budget constraint in Eq. (1), an external penalty function is used in the upper-level formulation. Therefore, it is rewritten as follows:

$$Z_k = \left\{\sum_{k\in K} t_k[v_k(y,\mu)]v_k(y,\mu)\right\} + \lambda * \left\{\text{Max}\left[0, \sum_{k\in K} g_k(\mu_k) - B\right]\right\} \tag{4}$$

where $Z_k$ = unconstrained objective function; $\lambda$ = penalty parameter; and $\sum_{k\in K} g_k(\mu_k) - B$ = constraint violation expression. The objective function $Z_k$, which consists of total system travel time and any possible penalty cost due to the total budget used that may exceed what is allowed, is passed to the GA procedure to compute the fitness values for each population and generation.

## Solution Methodology

### GA Implementation

Genetic algorithms, as a well-known type of adaptive heuristic search algorithms, are inspired by the evolutionary ideas of natural selection and genetics (Holland 1975; Goldberg 1989; Michalewicz 1999). Genetic algorithms have been proven to provide a robust search as well as a near-optimal solution in a reasonable amount of time. The working process of genetic algorithms is simple to understand; it involves nothing more than copying strings or swapping partial strings. The simplicity of the operations and the ability to find good solutions are two characteristics that make this method very suitable for solving network design problems (Fan and Machemehl 2006, 2011; Fan 2004).

### Genetic Algorithm Solution Procedure

This section presents a GA-based solution methodology to solve the optimal congestion pricing and continuous capacity investment problem. There are four decision variables involved in the GA procedure: toll charge, toll location, percentage capacity expansion, and the corresponding location. The number of bits for each variable needs to be determined first to form a chromosome of their combination that represents a possible optimal solution. For example, suppose a toll rate in a single link varies between \$2 and \$10 and assume a network consisting of 76 links, the toll locations, and toll levels are represented by seven bits and 10 bits, respectively, assuming the number of significant figures (i.e., the number of decimal points) for the toll rate is 2. Likewise, if percentage capacity expansion in a single link varies between 0 and 50%, it can be represented by six bits if the same number of decimal points is to be achieved. Detailed calculation can be found elsewhere (Fan and Machemehl 2006, 2011; Fan 2004). Fig. 1 shows chromosome structures for this example for the first six populations when the number of links selected for imposing links and capacity expansion are both one.

Fig. 2 presents a flow chart showing the whole GA process for the stated problem. The process starts by setting the number of toll and capacity expansion links equal to one. A possible set of solutions is initialized randomly. Each solution is evaluated based on the fitness values (the objective function values). The fittest solutions will then be selected for parenthood to perform crossover (using one-point crossover method) and mutation operations. Once the convergence criterion (the maximum number of generation in this case) is met, the number of links to be tolled will be set to two and the steps are repeated. It can be seen from Fig. 2 that there are also intermediate steps within the genetic algorithm process. For example, a lower-level network analysis has to be performed to evaluate the fitness values discussed previously. Lower-level network analysis involves determining the flows that are going to be used as inputs in the upper-level program using the most commonly applied Frank-Wolfe (FW) algorithm. The algorithm works iteratively by using an adaptive step size to calculate the right amount of flow to shift to get as close to equilibrium condition as possible. The detail algorithm is available elsewhere (Sheffi 1984).

| Toll Location | | | | | | | Toll Level | | | | | | | | | | Capacity Location | | | | | | | Capacity Expansion | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |

**Fig. 1.** Proposed chromosome structure

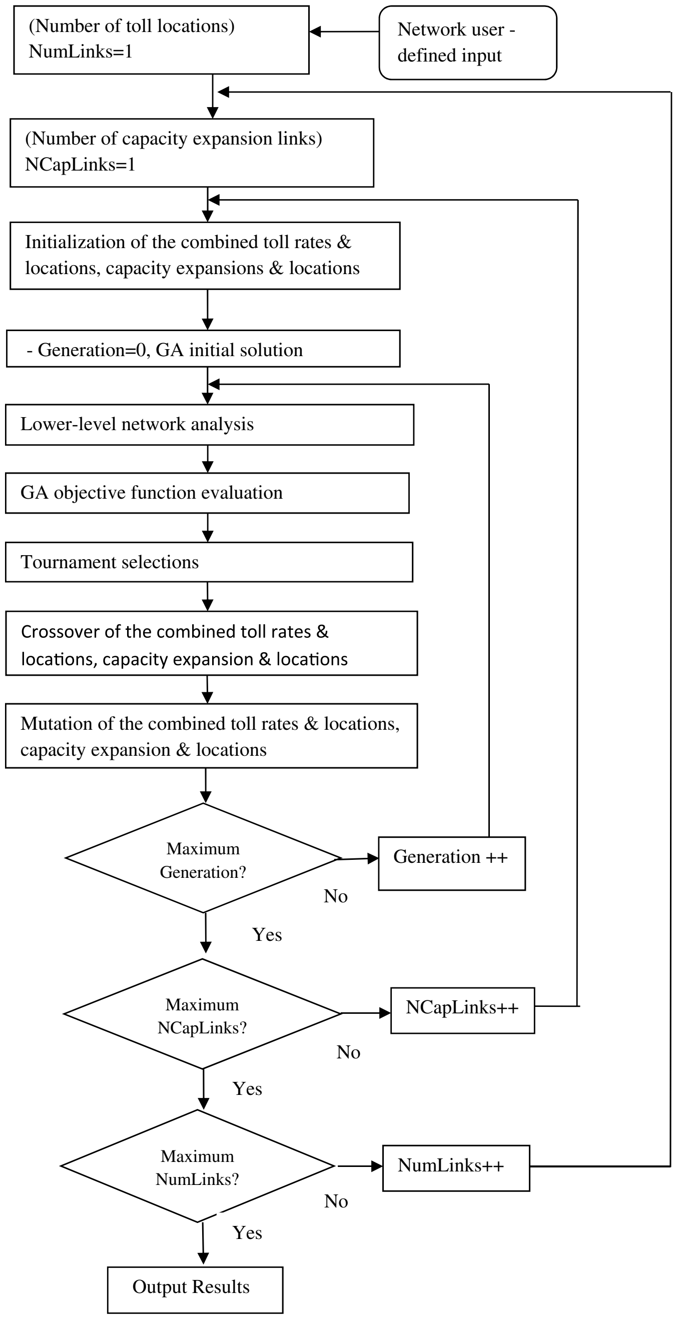


**Fig. 2.** Flow chart showing the GA procedure

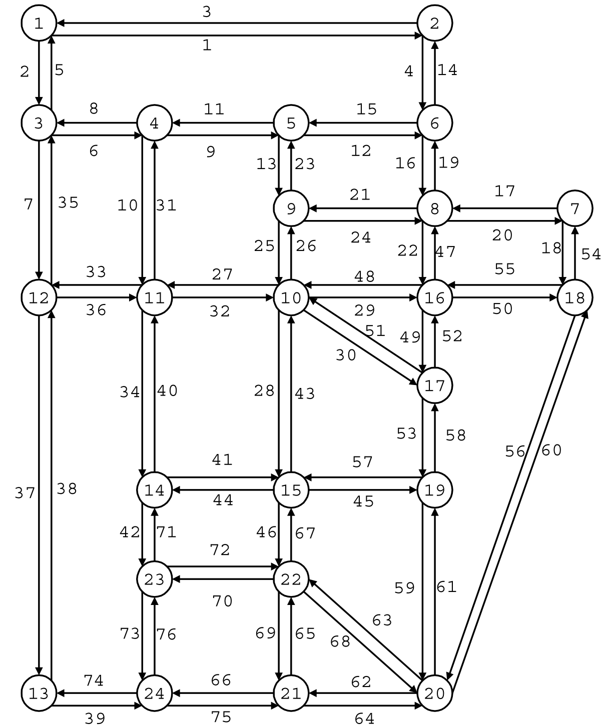

**Fig. 3.** Sioux Falls test network (data from Bar-Gera 2010)

## Numerical Experimentation

### Example Network Description

The Sioux Falls network shown in Fig. 3 is considered. This example network contains 24 travel demand zones, 76 links, and 576 O-D pairs (out of which 24 intrazonal and 24 interzonal zero-demand O-D pairs are deleted). This network has been used in many publications, as it is good for code debugging. Besides, Bar-Gera (2010) found the Sioux Falls user equilibrium solution using the quadratic BPR cost functions, which could be used for cross-checking results. Bar-Gera (2010) took all the network data, which are also used in this paper, from LeBlanc et al. (1975).

## Numerical Results

In this section, numerical experiments on the combined congestion pricing and continuous capacity expansion problem have been discussed using the Sioux Falls network with homogenous road users. The main objective here is to find the optimal toll rates, toll locations, percentage capacity expansion, and capacity expansion locations simultaneously. In practice, policy makers may first decide on the desired number of links where toll is to be charged and capacity is to be expanded. Normally, the most congested links could be chosen as candidate set of toll links. This set, however, may or may not give the optimal solution. Therefore, different combinations of number of links for toll charge and capacity expansion need to be investigated using the proposed GA solution methodology.

Parameters inherent in the GA algorithms need to be carefully chosen, as they will have some effects on the optimal solution. Based on some previous research efforts (Chen and Yang 1992; Recker et al. 2005), the following parameters are assumed in this numerical experimentation: population size, 64; maximum number of generation, 250; crossover probability, 0.7; mutation probability, 0.05; maximum toll rate ($y_k^{\max}$), \$10; minimum toll rate ($y_k^{\min}$), \$2; value of time (VOT), \$10/h; minimum capacity expansion ($\mu_k^{\min}$), 0%; maximum capacity expansion ($\mu_k^{\max}$), 20%; available budget for expansion ($B$), \$15 million; and penalty parameter ($\lambda$), 10,000. In addition, it is assumed that policy makers decided that a maximum of five links can be tolled and a maximum of five links can be expanded for the given network. This means there are 25 combinations to investigate, out of which one is the optimal solution.

Based on the preceding assumptions, the proposed GA procedure was implemented using *MATLAB* software package. The GA solution for all of the 25 possible combinations is presented in Table 1.

As can be seen from Table 1, the second combination, where the number of toll links is one and the number of links to be

**Table 1.** Toll Locations, Corresponding Rates, Capacity Locations and the Corresponding Percentage Capacity Expansions for all Combinations

| Number of toll links | Number of capacity links | Toll locations | | | | | Corresponding toll rates | | | | | Capacity expansion locations | | | | | Corresponding percentage capacity expansions | | | | | Budget (million $) | Total system travel time | Implementation benefits compared with base case | Tolling without capacity expansion | Capacity expansion without tolling |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 39 | | | | | 2.73 | | | | | 19 | | | | | 0.20 | | | | | 6.40 | 7,330,814.08 | 149,667.72 | 7,463,788.50 | 7,354,445.39 |
| 1 | 2 | 39 | | | | | 2.54 | | | | | 19 | 16 | | | | 0.20 | 0.17 | | | | 11.84 | 7,231,085.19 | 249,396.61 | 7,468,380.35 | 7,257,084.92 |
| 1 | 3 | 74 | | | | | 2.04 | | | | | 46 | 19 | 16 | | | 0.12 | 0.16 | 0.12 | | | 14.72 | 7,257,646.92 | 222,834.88 | 7,482,684.59 | 7,257,879.62 |
| 1 | 4 | 74 | | | | | 2.28 | | | | | 48 | 16 | 19 | 61 | | 0.06 | 0.10 | 0.20 | 0.02 | | 13.44 | 7,236,378.73 | 244,103.07 | 7,477,874.62 | 7,256,471.93 |
| 1 | 5 | 39 | | | | | 2.12 | | | | | 52 | 50 | 19 | 16 | 49 | 0.07 | 0.03 | 0.13 | 0.14 | 0.03 | 13.28 | 7,268,172.11 | 212,309.69 | 7,482,674.64 | 7,271,461.88 |
| 2 | 1 | 39 | 58 | | | | 2.75 | 2.07 | | | | 16 | | | | | 0.20 | | | | | 6.40 | 7,3289,98.24 | 151,483.56 | 7,469,294.92 | 7,360,102.88 |
| 2 | 2 | 51 | 19 | | | | 2.17 | 2.11 | | | | 13 | 16 | | | | 0.10 | 0.17 | | | | 13.44 | 7,338,595.71 | 141,886.09 | 7,477,275.69 | 7,344,007.78 |
| 2 | 3 | 52 | 53 | | | | 2.03 | 2.27 | | | | 16 | 39 | 34 | | | 0.18 | 0.10 | 0.01 | | | 12.80 | 7,296,518.37 | 183,963.43 | 7,462,429.26 | 7,319,699.21 |
| 2 | 4 | 48 | 16 | | | | 2.64 | 2.00 | | | | 49 | 19 | 52 | 9 | | 0.12 | 0.13 | 0.12 | 0.03 | | 12.80 | 7,313,936.98 | 166,544.82 | 7,485,481.15 | 7,327,414.41 |
| 2 | 5 | 39 | 74 | | | | 2.34 | 2.76 | | | | 16 | 65 | 13 | 19 | 25 | 0.17 | 0.10 | 0.00 | 0.15 | 0.03 | 14.88 | 7,249,805.23 | 230,676.57 | 7,463,552.29 | 7,275,786.43 |
| 3 | 1 | 29 | 48 | 74 | | | 2.92 | 2.80 | 2.74 | | | 19 | | | | | 0.19 | | | | | 6.08 | 7,320,497.78 | 159,984.02 | 7,457,918.68 | 7,352,967.29 |
| 3 | 2 | 51 | 48 | 19 | | | 2.20 | 2.93 | 2.55 | | | 16 | 28 | | | | 0.17 | 0.09 | | | | 14.08 | 7,328,083.12 | 152,398.68 | 7,489,607.08 | 7,312,583.90 |
| 3 | 3 | 74 | 49 | 58 | | | 2.21 | 2.20 | 2.01 | | | 44 | 16 | 20 | | | 0.01 | 0.12 | 0.08 | | | 8.48 | 7,359,646.07 | 120,835.73 | 7,446,489.20 | 7,384,448.08 |
| 3 | 4 | 5 | 14 | 29 | | | 4.04 | 4.23 | 3.02 | | | 48 | 10 | 57 | 25 | | 0.13 | 0.03 | 0.03 | 0.04 | | 14.56 | 7,422,931.53 | 57,550.27 | 7,503,697.00 | 7,415,169.29 |
| 3 | 5 | 66 | 68 | 64 | | | 2.06 | 2.34 | 2.18 | | | 67 | 32 | 45 | 74 | 46 | 0.01 | 0.10 | 0.04 | 0.02 | 0.05 | 14.08 | 7,456,427.26 | 24,054.54 | 7,517,266.18 | 7,407,568.23 |
| 4 | 1 | 73 | 69 | 4 | 2 | | 2.02 | 2.10 | 3.74 | 3.29 | | 48 | | | | | 0.17 | | | | | 10.88 | 7,418,058.58 | 62,423.22 | 7,487,214.51 | 7,394,478.55 |
| 4 | 2 | 68 | 39 | 64 | 16 | | 2.34 | 2.26 | 2.12 | 2.09 | | 19 | 14 | | | | 0.19 | 0.06 | | | | 10.88 | 7,415,986.27 | 64,495.53 | 7,515,429.90 | 7,353,471.36 |
| 4 | 3 | 16 | 74 | 14 | 5 | | 2.34 | 2.74 | 2.70 | 2.34 | | 28 | 18 | 58 | | | 0.14 | 0.01 | 0.03 | | | 14.72 | 7,408,926.66 | 71,555.14 | 7,498,651.83 | 7,388,589.82 |
| 4 | 4 | 48 | 5 | 14 | 39 | | 2.60 | 8.16 | 9.34 | 2.14 | | 70 | 74 | 19 | 66 | | 0.04 | 0.06 | 0.10 | 0.06 | | 12.48 | 7,381,890.73 | 98,591.07 | 7,509,109.13 | 7,365,958.76 |
| 4 | 5 | 52 | 74 | 58 | 53 | | 2.77 | 3.17 | 2.01 | 3.43 | | 63 | 22 | 14 | 19 | 32 | 0.01 | 0.01 | 0.01 | 0.14 | 0.10 | 14.88 | 7,394,117.13 | 86,364.67 | 7,543,842.78 | 7,350,107.07 |
| 5 | 1 | 51 | 62 | 69 | 39 | 30 | 2.34 | 2.33 | 3.23 | 2.95 | 2.45 | 27 | | | | | 0.17 | | | | | 13.60 | 7,524,670.61 | −44,188.81 | 7,568,696.42 | 7,421,990.39 |
| 5 | 2 | 68 | 48 | 31 | 64 | 58 | 3.94 | 2.88 | 2.54 | 3.94 | 3.89 | 25 | 28 | | | | 0.19 | 0.06 | | | | 14.88 | 7,527,234.25 | −46,752.45 | 7,607,391.50 | 7,410,505.81 |
| 5 | 3 | 5 | 14 | 51 | 70 | 39 | 2.70 | 3.62 | 2.43 | 2.08 | 3.19 | 41 | 46 | 19 | | | 0.02 | 0.04 | 0.13 | | | 7.68 | 7,454,152.60 | 26,329.20 | 7,540,326.72 | 7,372,304.83 |
| 5 | 4 | 58 | 75 | 19 | 65 | 76 | 2.55 | 2.79 | 2.13 | 2.30 | 3.50 | 8 | 67 | 10 | 44 | | 0.05 | 0.10 | 0.01 | 0.03 | | 11.36 | 7,566,218.05 | −85,736.25 | 7,592,125.73 | 7,443,502.63 |
| 5 | 5 | 39 | 48 | 46 | 53 | 52 | 2.92 | 3.11 | 2.56 | 3.45 | 3.33 | 22 | 19 | 52 | 35 | 3 | 0.01 | 0.17 | 0.19 | 0.01 | 0.01 | 13.92 | 7,513,557.34 | −33,075.54 | 7,599,019.97 | 7,329,201.13 |



J. Transp. Eng.

expanded is two, gives the optimal solution, as it has the lowest total system travel cost value (i.e., 7,231,085.19). Because all budgets used are below $15 million (which is the assumed available budget for expansion) and there will be no penalty cost accrued, the total system travel cost [as shown in Eq. (4)] is also equal to the total system travel time, as used in Tables 1 and 2. As part of preliminary analysis, the total system costs for non-tolling equilibrium and system optimum cases before network improvement were found to be 7,480,481.8 and 7,246,945.8, respectively. One column labeled "Implementation benefits compared with base case" is also provided to show the benefits (i.e., the reduction in total system travel time) with improvements of implementing tolling, capacity expansion, combined or separately, compared with the total travel time without any improvement in Tables 1 and 2. Also, the numerical experimentation revealed that charging a single link and expanding two links gives a little better result than even the system optimum case (but without capacity additions), resulting in increased network performance. It can be observed from Tables 1 and 2 that it is not always advantageous to consider large number of toll or capacity links when improving a network. For example, the study showed that charging or expanding a large number of links may even be worse than not charging or expanding at all. In addition, increasing the capacity of more links may generate better network performance, as clearly indicated in Table 2. However, tolling seems to behave differently. As the number of links charged increases, the performance generally gets worse, as also shown in Table 2. When tolling and capacity expansion are combined, the general trend is that charging or expanding a larger number of links may produce worse performance, as can be seen in Table 1. This might be because adding more toll locations may make travelers shift their routes in an undesirable way so as to increase the total system travel costs.

Another observation is that the toll link, i.e., link 39, in the optimal solution, is not among the five most congested links. However, it appears that the GA procedure found link 16 and link 19, which are the two most congested links in the Sioux Falls network, as locations for capacity expansions, as one would expect. This may suggest that selecting the most congested links for capacity expansion might be an ideal solution to improve the network performance; however, charging at the most congested links may not work well for the system. For comparison purposes, two different but relevant strategies (including tolling without capacity expansion and capacity expansion without tolling) are also considered, and corresponding numerical results are presented in two columns (titled "Tolling without capacity expansion" and "Capacity expansion without tolling") by performing two separate optimization software runs, in which the optimized results listed in the first six columns are taken, and the toll locations and levels, or capacity expansions are removed, respectively, for each case. In particular, for each combination in Table 1, the combined tolling and capacity expansion strategy always outperforms the tolling without capacity expansion and capacity expansion without tolling strategies. However, further comparisons made also indicate that several capacity expansion only optimal solutions as presented in Table 2 work even better in terms of their lower total system travel costs/times than the combined tolling and capacity expansion optimal solution, as shown in Table 1. Nevertheless, further comprehensive investigations need to be conducted to confirm the hypothesis that the most congested links may be taken as candidate locations for capacity expansion by default, and that capacity expansion only may be considered in many cases, to achieve best network performance.

In addition, it is seen in Table 2 that adding more toll locations tends to initially decrease and then later increase the total system travel costs/times. This might suggest the importance of selecting the number of toll locations (in addition to toll rates), because adding more toll locations may make travelers shift their routes in an undesirable way so as to increase the total system travel costs. This is somewhat like Braess's paradox, in which adding extra capacity to a network on some links can, in some cases, reduce overall performance. Further comprehensive investigations need to be conducted to confirm such hypothesis.

Generally, GA terminates when either a maximum number of generations has been produced or the fitness value can no longer be improved for a certain number of successive iterations. In this study, the number of genetic algorithm iterations is used as stopping criteria. Therefore, it is important to conduct sensitivity analysis to examine how the total system travel cost behaves as the number of iterations increases. Fig. 4 presents the sensitivity of travel cost over different generations, for different number of capacity expansion locations, when only a single link is assumed to be tolled. As can be seen from the figure, the value of total system travel cost declined sharply at the very beginning and further declined but at a much slower rate until the convergence criteria is met for all number of links to be expanded. This illustrates that every

**Table 2.** Toll Only and Capacity Expansion Only

| Tolling only | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of toll links | Toll locations | | | | | Corresponding toll rates | | | | | Total system travel time | Implementation benefits compared with base case |
| 1 | 58 | | | | | 2.03 | | | | | 7,457,996.50 | 22,485.30 |
| 2 | 39 | 74 | | | | 2.77 | 2.47 | | | | 7,449,141.44 | 31,340.36 |
| 3 | 74 | 14 | 5 | | | 2.83 | 3.52 | 2.60 | | | 7,476,671.19 | 3,810.61 |
| 4 | 39 | 48 | 62 | 63 | | 2.51 | 2.84 | 2.50 | 2.88 | | 7,494,809.21 | −14,327.41 |
| 5 | 74 | 51 | 52 | 39 | 65 | 3.14 | 2.98 | 2.12 | 2.23 | 2.42 | 7,546,769.64 | −66,287.84 |

| Capacity expansion only | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of capacity links | Capacity expansion locations | | | | | Corresponding percentage capacity expansions | | | | | Total system travel time | Implementation benefits compared with base case | Budget (million $) |
| 1 | 19 | | | | | 0.19 | | | | | 7,352,967.29 | 127,514.51 | 6.08 |
| 2 | 19 | 16 | | | | 0.20 | 0.20 | | | | 7,224,523.20 | 255,958.60 | 12.80 |
| 3 | 16 | 24 | 19 | | | 0.20 | 0.01 | 0.20 | | | 7,203,060.54 | 277,421.26 | 14.40 |
| 4 | 16 | 19 | 54 | 49 | | 0.17 | 0.15 | 0.04 | 0.10 | | 7,241,318.51 | 239,163.29 | 14.72 |
| 5 | 29 | 16 | 19 | 70 | 55 | 0.01 | 0.18 | 0.20 | 0.01 | 0.02 | 7,220,228.48 | 260,253.32 | 14.40 |

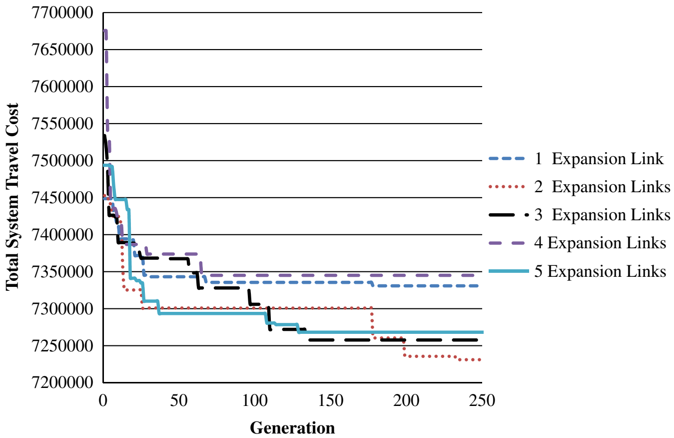


**Fig. 4.** Sensitivity of total system travel cost over generation and number of toll links

new generation gives results that are equal to or better than the previous ones, as might be expected.

It was found that charging a single link and expanding two links gave the optimal solution, assuming the available budget for expansion equal to \$15 million. However, it is important to conduct budget sensitivity analysis, as it helps planners estimate what will happen to the project if the budget is actually uncertain and subject to change given the uncertain economic conditions. The analysis involves changing the budget assumptions in a calculation to see the impact on the total system travel cost/time and decision variables. The number of toll links and capacity locations for expansion were fixed to be one and two, respectively, and the whole GA procedure was repeated for different values of budgets. Table 3 presents the results of the sensitivity analysis for different budgets.

It can be seen from Table 3 that budget is satisfied in all cases, as the required budget for network improvement is less than the corresponding available budget for each case. As the available budget increases, it is observed that the total system travel time for the network generally decreases, as one would anticipate. This may be because with a higher budget available, it is likely to achieve a higher percentage of capacity expansion, which typically will improve network performance by minimizing total system travel cost.

As discussed previously, the budget constraint is handled using a penalty function in the upper-level program. In the original analysis, a higher penalty parameter value ($\lambda = 10{,}000$) was assumed. With the budget fixed to \$15 million, sensitivity of the penalty parameter was conducted to see its impact on the network performance. Table 4 presents the results of the sensitivity analysis for the penalty parameter.

As can be seen from the Table 4, although the same level of network performance can be achieved with certain values of penalty parameters (e.g., $\lambda = 100$, 500, and 1,000), the total system travel time generally decreases as the value of the penalty parameter increases. Also, as the penalty parameter increases, the GA procedure finds the two most congested links with high percentage of capacity expansion. This might suggest that by setting a higher penalty, GA will behave more efficiently toward finding better solutions along the process, although no penalty is actually incurred for any set penalty parameter because all required budgets are satisfied. The number of generations used for this numerical experimentation is 250, and perhaps it is possible to get closer to the optimal results in the preceding sensitivity analyses if a larger number of iterations was considered instead.

**Table 3.** Budget Sensitivity Analysis

| Number of toll links | Number of capacity links | Toll locations | Corresponding toll rates | Capacity expansion locations | | Corresponding percentage capacity expansions | | Total system travel time | Implementation benefits compared with base case | Available budget (in million \$) | Required budget (in million \$) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 29 | 2.83 | 25 | 19 | 0.03 | 0.14 | 7,363,425.91 | 117,055.89 | 6 | 5.92 |
| 1 | 2 | 74 | 3.09 | 16 | 21 | 0.20 | 0.01 | 7,333,059.75 | 147,422.05 | 9 | 8.00 |
| 1 | 2 | 39 | 2.77 | 16 | 19 | 0.16 | 0.19 | 7,240,668.62 | 239,813.18 | 12 | 11.20 |
| 1 | 2 | 39 | 2.54 | 16 | 19 | 0.20 | 0.17 | 7,231,085.19 | 249,396.61 | 15 | 11.84 |
| 1 | 2 | 74 | 2.43 | 19 | 39 | 0.20 | 0.15 | 7,229,562.16 | 250,919.64 | 18 | 16.00 |
| 1 | 2 | 58 | 2.11 | 19 | 39 | 0.20 | 0.19 | 7,216,078.30 | 264,403.50 | 21 | 18.56 |
| 1 | 2 | 39 | 2.38 | 39 | 19 | 0.20 | 0.20 | 7,205,292.95 | 275,188.85 | 24 | 19.20 |

**Table 4.** Penalty Parameter Sensitivity Analysis

| Number of toll links | Number of capacity links | Toll locations | Corresponding toll rates | Capacity expansion locations | Corresponding percentage capacity expansions | Total system travel time | Implementation benefits compared with base case | Penalty parameter | Required budget (in million $) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 74 | 2.08 | 48 16 | 0.15 0.20 | 7,314,356.53 | 166,125.27 | 1 | 14.40 |
| 1 | 2 | 39 | 2.60 | 32 16 | 0.10 0.19 | 7,298,322.49 | 182,159.31 | 50 | 14.08 |
| 1 | 2 | 39 | 2.47 | 48 19 | 0.12 0.19 | 7,295,268.44 | 185,213.36 | 100 | 13.76 |
| 1 | 2 | 39 | 2.47 | 48 19 | 0.12 0.19 | 7,295,268.44 | 185,213.36 | 500 | 13.76 |
| 1 | 2 | 39 | 2.47 | 48 19 | 0.12 0.19 | 7,295,268.44 | 185,213.36 | 1,000 | 13.76 |
| 1 | 2 | 39 | 2.58 | 19 16 | 0.18 0.19 | 7,240,702.75 | 239,779.05 | 5,000 | 11.84 |
| 1 | 2 | 39 | 2.54 | 16 19 | 0.20 0.17 | 7,231,085.19 | 249,396.61 | 10,000 | 11.84 |

## Summary and Future Research

This paper attempts to develop a GA-based bilevel optimization solution methodology to determine the optimal toll location, toll rate, percentage capacity expansion, and the location for expansion simultaneously in a well-known example network for which the demand is assumed to be fixed and given a priori. Numerical experiments were conducted using Sioux Falls network by assuming an available budget of $15 million for expansion. The optimal solution for the stated problem was found when two links are considered for capacity expansion and a single link for toll charging. The two locations for capacity expansion were found to be the two most congested links in the network. However, it appears that the single link for toll charging is not even among the five most congested links. However, further numerical experiments need to be conducted to confirm if these findings are always true. Sensitivity analysis for budget was conducted as it helps planners estimate the impact on the capacity improvement project if the budget is actually uncertain and subject to change. The sensitivity analyses revealed that, as the available budget increases, the total system travel time for the network generally decreases. This may be because the higher the budget, the higher the percentage of capacity expansion. Sensitivity analysis for the penalty parameter was also conducted. It was found that the required budget for expansion can be minimized by setting a higher penalty parameter value. In conclusion, the proposed methodology will be a very useful tool for transportation network planners for allocation of budgets and prioritization of links for improvements and congestion pricing.

However, the variation of demand across different times of the day was not considered. Future research may be directed toward this end with further insight provided for solving combined capacity investment and congestion charging problem by considering stochastic dynamic demand, as it is a very important issue in both design of new and redesign of existing road networks. In addition, the costs of setting up and operating toll strategies, dynamic (instead of fixed) toll rates, discrete toll rate (for example, the toll rate should always be rounded to the closest $0.25 for most real-world applications) as opposed to the continuous toll rate in this paper, as well as the maximization of toll revenue in the upper-level objective function (rather than the minimization of the total system travel cost) may be considered in the future. Additional interesting insights to the model comparisons and numerical results may be presented accordingly. Furthermore, in the optimization model developed in this paper, both toll locations and capacity expansion locations are unconstrained; that is, any link can be a candidate for toll and/or capacity expansion. In reality, however, due to the geometry and accessibility limitation, it is very likely that only a subset of links is tollable. Similarly, most capacity expansion projects are usually selected for implementation at bottleneck locations to reduce traffic congestion. Therefore, for real-world applications, perhaps future research should consider adding such location constraints in the optimization model, which can limit the optimal solutions to meaningful results and also reduce the running time of optimization process. Last but not least, future research directions will be directed toward developing global optimization techniques for solving congestion pricing and/or capacity expansion.

## Notation

*The following symbols are used in this paper:*

$B$ = available budget for expansion;

$f_p^w$ = flow on path $p \in P_w$ between O-D pair $w \in W$;

$K$ = Set of links (arcs) such that $k \in K$;

$\bar{K}$ = subset of links to be tolled i.e., $\bar{K} \subseteq K$;

$\overline{\overline{K}}$ = subset of links for capacity expansion i.e., $\overline{\overline{K}} \subseteq K$, $\overline{\overline{K}} \cap \bar{K} = \emptyset$;
$k$ = Link;
$N$ = number of nodes;
$n$ = Node;
$P_w$ = Set of paths between O-D pair $w \in W$;
$q_w$ = a priori demand between O-D pair w;
$t_k(v_k)$ = travel time on link $k \in K$ given $v_k$;
VOT = average value of time;
$v_k$ = the link flow on link $k \in K$;
$W$ = Set of O-D pairs;
$w$ = Origin-Destination (O-D) pair;
$y_k$ = toll level on link $k \in \bar{K}$;
$y_k^{\max}$ = upper-bound toll level of link $k \in K$;
$y_k^{\min}$ = lower-bound toll level of link $k \in K$;
$\delta_{kp}^{w} = 1$ = if link $k$ is used in path $p$ or $\delta_{kp}^{w} = 0$ otherwise, $w \in W$;
$\mu_k$ = capacity expansion i.e., $k \in \overline{\overline{K}}$;
$\mu_k^{\max}$ = upper-bound capacity expansion of link $k \in K$; and
$\mu_k^{\min}$ = lower-bound capacity expansion of link $k \in K$.

## References


Bar-Gera, H. (2010). "Transportation test problems." ⟨http://www.bgu.ac.il/~bargera/tntp/⟩ (Apr. 16, 2013).

Bar-Gera, H., Hellman, F., and Patriksson, M. (2013). "Computational precision of traffic equilibria sensitivities in automatic network design and road pricing." *Transp. Res. Part B*, 57, 485–500.

Chen, A., and Yang, C. (1992). "Stochastic transportation network design problem with spatial equity constraint." *Transportation Research Record 1882*, Transportation Research Board, Washington, DC, 97–104.

Clegg, J., Smith, M., Xiang, Y., and Yarrow, R. (2001). "Bi-level programming applied to optimizing urban transportation." *Transp. Res. Part B*, 35(1), 41–70.

Cree, N. D., Maher, M. J., and Paechter, B. (1998). "The continuous equilibrium optimal network design problem: A genetic approach." *Transportation networks: Recent methodological advances*, Pergamon, Amsterdam, 163–174.

Ekström, J., Sumalee, A., and Lo, H. K. (2012). "Optimizing toll locations and levels using a mixed integer linear approximation approach." *Transp. Res. Part B*, 46(7), 834–854.

Fan, W. (2004). "Optimal transit route network design problem: Algorithms, implementations, and numerical results." Ph.D. dissertation, Univ. of Texas at Austin, Austin, TX.

Fan, W., and Machemehl, R. B. (2006). "Optimal transit route network design problem with variable transit demand: A genetic algorithm approach." *J. Transp. Eng.*, 10.1061/(ASCE)0733-947X(2006)132:1(40), 40–51.

Fan, W., and Machemehl, R. B. (2011). "Bi-level optimization model for public transportation network redesign problem: Accounting for equity issues." *Transportation Research Record 2263*, Transportation Research Board, Washington, DC, 151–162.

Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., and Rashidi, H. (2013). "A review of urban transportation network design problems." *Eur. J. Oper. Res.*, 229(2), 281–302.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, London.

Hearn, D. W., and Ramana, M. V. (1998). "Solving congestion toll pricing models." *Equilibrium and Advanced Transportation Modelling*, Springer, New York.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor, MI.

LeBlanc, L. J., Morlok, E. K., and Pierskalla, W. P. (1975). "An efficient approach to solving the road network equilibrium traffic assignment problem." *Transp. Res.*, 9(5), 309–318.

Li, C. M., Yang, H., Zhu, D. L., and Meng, Q. (2012). "A global optimization method for continuous network design problems." *Transp. Res. Part B*, 46(9), 1144–1158.

Lo, H., and Szeto, W. Y. (2003). "Time-dependent transport network design: A study of budget sensitivity." *J. East. Asia Soc. Transp. Stud.*, 5, 1124–1139.

Mathew, T. V., and Sharma, S. (2009). "Capacity expansion problem for large urban transportation networks." *J. Transp. Eng.*, 10.1061/(ASCE)0733-947X(2009)135:7(406), 406–415.

Miandoabchi, E., Daneshzand, F., Szeto, W. Y., and Farahani, R. Z. (2013). "Multi-objective discrete urban road network design." *Comput. Oper. Res.*, 40(10), 2429–2449.

Miandoabchi, E., and Farahani, R. Z. (2011). "Optimizing reserve capacity of urban road networks in a discrete network design problem." *Adv. Eng. Softw.*, 42(12), 1041–1050.

Michalewicz, Z. (1999). *Genetic algorithms + Data structure = Evolution programs*, 3rd Ed., Springer, New York.

Ordonez, F., and Zhao, J. (2007). "Robust capacity expansion of network flows." *Networks*, 50(2), 136–145.

Recker, W., et al. (2005). *Considering risk-taking behavior in travel time reliability*, Institute of Transportation Studies, Univ. of California, Irvine, CA.

Sheffi, Y. (1984). *Urban transportation networks: Equilibrium analysis with mathematical programming methods*, Prentice-Hall, Englewood Cliffs, NJ.

Shepherd, S., and Sumalee, A. (2004). "A genetic algorithm based approach to optimal toll level and locations problems." *Network Spat. Econ.*, 4(2), 161–179.

Verhoef, E. T. (2002). "Second-best congestion pricing in general networks: Heuristic algorithms for finding second-best optimal toll levels and toll points." *Transp. Res. Part B*, 36(8), 707–729.

Verhoef, E. T., Nijkamp, P., and Ritveld, P. (1996). "Second-best congestion pricing: The case of an untolled alternative." *J. Urban Econ.*, 40(3), 279–302.

Wang, S., Meng, Q., and Yang, H. (2013). "Global optimization algorithms for the discrete network design problem." *Transp. Res. Part B*, 50, 42–60.

Yang, H. (1996). "Equilibrium network traffic signal setting under conditions of queuing and congestion." *4th Applications of Advanced Technologies in Transportation Engineering Int. Conf.*, Capri, Italy, 578–582.

Yang, H., and Bell, M. G. H. (1997). "Traffic restraint, road pricing and network equilibrium." *Transp. Res. Part B*, 31(4), 303–314.

Yang, H., and Bell, M. G. H. (1998). "Models and algorithms for road network design: A review and some new development." *Transp. Rev.*, 18(3), 257–278.

Yang, H., and Huang, H. J. (2005). *Mathematical and economic theory of road pricing*, Emerald Group Publishing Limited, Bradford, U.K.

Yang, H., and Zhang, X. (2003). "Optimal toll design in second-best link-based congestion pricing." *Transp. Res. Rec.*, 1857, 85–92.

Zhang, H., and Gao, Z. (2009). "Bi-level programming model and solution method for mixed transportation network design problem." *J. Syst. Sci. Complex.*, 22, 446–459.

Zhang, L., and Sun, J. (2013). "Dual-based heuristic for optimal cordon pricing design."*J. Transp. Eng.*, 10.1061/(ASCE)TE.1943-5436.0000591, 1105–1116.

Zhang, X., and Yang, H. (2004). "The optimal cordon-based network congestion pricing problem." *Transp. Res. Part B*, 38(6), 517–537.