



ROAD NETWORK RESERVE CAPACITY WITH STOCHASTIC USER EQUILIBRIUM

Jian Wang^{1,2}, Wei Deng¹, Jinbao Zhao¹

¹*School of Transportation, Southeast University, China*

²*School of Civil Engineering, Purdue University, United States*

Submitted 7 May 2013; resubmitted 10 July 2013, 20 September 2013, 27 November 2013;
accepted 14 December 2013

Abstract. To relax the strong assumption associated with User Equilibrium (UE) in the previous research of network reserve capacity conducted by Gao and Song (2002), this paper assumes that the drivers all make route choices based on Stochastic User Equilibrium (SUE) principle. Similarly, two bi-level programs are formulated to study the network reserve capacity with SUE problem. The first bi-level program is developed to maximize the network reserve capacity by optimizing signal settings while the traffic demands are reassigned by SUE model. The second program extends the research with Continuous Network Design (CND) problem to find the maximum possible increase in reserve capacity through optimizing allocation of network investment. Two methods, i.e. the sensitivity analysis-based method and Genetic Algorithm (GA), are detailed formulated to solve the bi-level reserve capacity problem. Application of the proposed model and its solution algorithms on two numerical examples find that the network reserve capacity does not always increase with improved quality of drivers' information. Besides, CND can not only help to increase network reserve capacity, but also can help to make more use of physical capacity of road network at Deterministic User Equilibrium (DUE) condition, thus reduces the difference of reserve capacity between the assumptions of SUE and DUE.

Keywords: reserve capacity; stochastic user equilibrium; sensitivity analysis; genetic algorithm; continuous network design.

Introduction

Network reserve capacity is defined as the maximum additional demand that can be accommodated by a road network without exceeding a prescribed degree of saturation while taking users' route choice into account (Wong, Yang 1997). Akin to system travel time, reserve capacity is usually taken as an important performance indicator in network design and planning, thus the corresponding research is quite necessary and meaningful (Ge *et al.* 2003). One significant application for reserve capacity is to relieve traffic congestion in inner cities, by determining optimum input parameters such as the signal timings (circle time, splits, and so forth) link capacity increase and so on, the network can accommodate a maximum possible increase in travel demands. Besides, as an indicator represents the state of network, it is usually taken as a significant quantified guideline in estimating the effectiveness of some improvement measures.

Webster and Cobbe (1966) firstly defined the concept of reserve capacity of an intersection. Wong and Yang (1997) extended it into a general signal-controlled

road network under time-stationary conditions with Deterministic User Equilibrium (DUE) problem, a bi-level programming model was formulated and heuristic algorithm based on sensitivity analysis method was proposed to solve the problem. Gao and Song (2002) redefined the concept of reserve capacity by assuming that the demand multipliers between each Origin–Destination (O–D) pair could be different, better results were reported from the same example networks designed by Wong and Yang (1997). Chiou (2007) formulated a mathematical program with equilibrium constraints to maximize the reserve capacity with optimal signal settings. A projected gradient approach based on the Traffic Network Study Tool (TRANSYT) (Vincent *et al.* 1980) was proposed to solve the bi-level program. Furthermore, through embedding the concept of reserve capacity of signal-controlled intersection, Chiou (2008) formulated a bi-level optimization program in which the travel demand is maximized while the total cost for link capacity expansions and incurred delays of travelers are minimized at the same time. Yang *et al.* (2000) put forward a bi-level program to model the network capacity



with level of service problem. In this model, the drivers are allowed to have both routes and destination choices based on travel costs and the destination attractiveness measures. Subsequently, the concept of reserve capacity has been further utilized to compare the magnitude of network capacity under different definitions (Kasikitwiwat, Chen 2005; Chen, Kasikitwiwat 2011). Current literatures from Zhang *et al.* (2010), Miandoabchi and Farahani (2011) studied the reserve capacity with Discrete Network Design (DND) problem which aims to maximize network reserve capacity through optimizing configuration of street directions and lane allocations.

Most of the literatures mentioned above studied the reserve capacity problem at DUE condition, in which the drivers are assumed to have perfect knowledge of network travel cost and make route choice based on Wardrop's first principle. However, this assumption is far from practical that the quality of driver information is generally limited in real situations (Prashker, Bekhor 2004). Chen *et al.* (1999, 2002) extended the research with probit-based Stochastic User Equilibrium (SUE) problem. The concept of network capacity reliability was introduced to study the probability of a certain level of traffic demand that the road network can accommodate at equilibrium conditions. A Monte-Carlo simulation procedure was developed to estimate the capacity related reliability measure. Chen *et al.* (2006) further modeled a bi-level programming to study a new reserve capacity problem. Different from the integrated model formulated by Wong and Yang (1997), the up-level program for the new reserve capacity problem is to maximize the capacity reliability index instead of traditional O-D multipliers. Wang and Deng (2013) studied the reserve capacity on multi-phase signalized road network, an equivalent program for logit-based SUE problems was formulated to describe drivers' route choice on multi-phase signalized road network. Ceylan and Bell (2004) proposed a two-stage approach for reserve capacity problem, the first stage was to optimize signal timing to maximize the network performance while traffic was reassigned by logit-based SUE, and the second stage was to find the largest common multiplier at the optimized signal timings provided by the first stage. Ge *et al.* (2003) proposed a hierarchical framework to study reserve capacity with SUE constraints. A Genetic Algorithm (GA) based solution method was developed to solve the reserve capacity problem at different quality of traveler information.

To relax the assumptions associated with User Equilibrium (UE) model in previous reserve capacity research of Gao and Song (2002). This paper, we assume that the drivers all make route choice based on logit-based SUE principle. With this assumption, the drivers may have more diversity route choices according to their perceived travel utility, of which the accuracy is significantly depend on the quality of traveler information. Similarly, two models are proposed, one is the reserve capacity with SUE problem, and the other combines reserve capacity with Continuous Network Design (CND) problem. With the proposed sensitivity analysis method

for logit-based SUE model, a gradient-based method is explicitly developed to solve the two bi-level reserve capacity problems. Besides, to test the efficiency and effectiveness of the new method, GA which is constantly used for bi-level problems is also presented to make a comparison.

The remainder of this paper is structured as follows: Followed the introduction section, some basic notations used in this paper are defined. In Section 2, the most popular discrete route choice model is presented, and then two integrated methods are proposed to maximize the network reserve capacity. In Section 3, two algorithms: Sensitivity Analysis Based (SAB) method and GA are introduced to solve the bi-level problem. Section 4 presents a numerical example to illustrate the general use of proposed method for reserve capacity problem and comparison is conducted between DUE and SUE. Detailed discussions are also given regarding to the difference of reserve capacity with different quality of travelers' information. The last section concludes the paper.

1. Basic Notations

- N – the set of nodes in the network;
- A – the set of links in the network;
- \bar{A} – the set of all signal-controlled links in the network;
- a – a link in the network, $a \in A$;
- r – an origin node, $r \in N$;
- s – a destination node, $s \in N$;
- c_k^{rs} – the route travel cost on k th route between O-D pair $r-s$;
- P_k^{rs} – probability that a traveler from r to s choose path k ;
- R_{rs} – the set of all routes between O-D pair $r-s$;
- $s_a(\lambda_a, \gamma_a)$ – the capacity of signal-controlled link a , is a function of signal splits λ_a and γ_a ;
- q_{rs} – the O-D demand between O-D pair $r-s$;
- q_{rs}^0 – initial O-D demand between O-D pair $r-s$;
- v_a – the link flow on link a ;
- \mathbf{v} – a vector of all link flows;
- t_a – the travel cost on link a ;
- μ_{rs} – O-D multiplier of O-D pair $r-s$, the multiplied O-D demand is $\mu_{rs} q_{rs}^0$;
- $\boldsymbol{\mu}$ – a vector of all O-D matrix multipliers;
- λ_a – signal splits of link a ;
- $\boldsymbol{\lambda}$ – a vector of all signal splits;
- γ_a – the capacity increase on link $a \in A$;
- \mathbf{y} – a vector of all capacity increase;
- J – the set of all signalized intersections in the network;
- j – an signalized intersection on the road network, $j \in M$;
- $G_a(\gamma_a)$ – the investment function of link a , $a \in A$;
- TI – total investment for link improvement;
- $\delta_{a,k}^{rs}$ – link-route indicators, $\delta_{a,k}^{rs} = 1$ if a is a link on route k , else $\delta_{a,k}^{rs} = 0$;
- $\delta_{b,k}^{rs}$ – link-route indicators, $\delta_{b,k}^{rs} = 1$ if a is a link on route k , else $\delta_{b,k}^{rs} = 0$;
- θ – dispersion parameter in SUE.

2. Bi-Level Programming for Reserve Capacity with SUE Problem

2.1. Logit-Based SUE Problem

SUE assumes that the drivers always make route choice to minimize their perceived utility. Here, utility may be expressed as a random variable consisting of a deterministic component V_k and a random error component ξ_k , which can be regarded as the unobservable or unmeasurable factors (Sheffi 1985). For a given road network $G=(N,A)$. The individual utility is given by:

$$U_k^{rs} = V_k^{rs} + \xi_k^{rs}, k \in R_{rs}. \quad (1)$$

It is generally assumed that among many other attributes, the route travel cost c affects drivers' route choice mostly. Then the deterministic in equation (1) could be expressed as:

$$V_k = -c_k, k \in R_{rs}. \quad (2)$$

Consequently, to a single traveler, the utility received from one trip between O-D pair $r-s$ is formulated as:

$$U_k^{rs} = -c_k^{rs} + \xi_k^{rs}, k \in R_{rs}. \quad (3)$$

According to random utility theory (Prashker, Bekhor 2004), each traveler tries to maximize the utility U_k^{rs} of choosing route k from the set of routes. Then the probability of travelers choosing route k is given by:

$$P_k^{rs} = \text{prob}(U_k^{rs} \geq U_r^{rs}, \forall r \in R_{rs}), k \in R_{rs}. \quad (4)$$

It could be seen from equation (3) that the stochastic of perceived travel costs is accounted for by random variable ξ_k^{rs} . Different distributions of random variables will lead to different structure of the choice models. Generally, drivers' perception error ξ_k^{rs} is assumed to be Normal distribution (probit models) and Gumbel distribution (logit models). It is worth to note that although probit-based models are commonly believed to be more preferable than logit-based ones, which takes no account of overlapping, or correlated routes (Maher, Hughes 1997), it is less used and studied compared with logit-based SUE due to the lack of a closed-form probability expression. Therefore, in this paper, we assume that the random variables in equation (1) are identically and independently distributed Gumbel variables (Damberg *et al.* 1996), which resulting in logit-type SUE models, the corresponding route choice probabilities are specified as:

$$P_k^{rs} = \frac{\exp(-\theta c_k^{rs})}{\sum_{r \in R_{rs}} \exp(-\theta c_r^{rs})}, k \in R_{rs}. \quad (5)$$

Fisk (1980) formulated a well-known equivalent optimization program for logit-type SUE. However, for convenience of sensitivity analysis, the unconstrained mathematical program proposed by Sheffi and Powell (1982) is employed in this paper, whose minimized flow solutions satisfy both the logit-type and the probit-type SUE principle, the program is formulated as:

$$\min Z = \sum_{a \in A} v_a t_a(v_a) - \sum_{a \in A} \int_0^{v_a} t_a(x) dx - \sum_{\forall r,s} q_{rs} S_{rs}(\mathbf{c}^{rs}(\mathbf{v})), \quad (6)$$

where: $S_{rs}(\mathbf{c}^{rs}(\mathbf{v}))$ is the satisfaction function, defined as:

$$S_{rs}(\mathbf{c}^{rs}(\mathbf{v})) = E \left[\min_{k \in R_{rs}} (c_k^{rs}) | \mathbf{c}^{rs}(\mathbf{v}) \right] = -\frac{1}{\theta} \ln \sum_k \exp(-\theta c_k^{rs}). \quad (7)$$

It should be noted that although the flow pattern that minimize the program (6) coincides with the SUE conditions and is unique, it cannot be easily obtained by directly minimizing the program (6) because of uncertainty term of satisfaction function $S_{rs}(\mathbf{c}^{rs}(\mathbf{v}))$, which makes it difficult in searching descent direction on the iterative process. Generally the Method of Successive Averages (MSA) proposed by Powell and Sheffi (1982) are used. The main feature of the method is the step size determination: it is based on a decreasing function with respect to the number of iterations and hence is not dependent on the objective function (Prashker, Bekhor 2004). Another effective method for implementing a logit route choice model at the network level is the Dial's algorithm (Dial 1971). This algorithm only assigns choice probabilities to the routes that is reasonable and excludes routes that would not be considered in practice, thus reduce the computational complexity greatly.

2.2. Bi-level Programming for Reserve Capacity with SUE Problem

In equilibrium condition, the link flow is perturbed by demand multipliers vector μ , signal timing variables vector λ , and link capacity increase vector γ . Therefore, link flow could be regarded as a function of the O-D demand multipliers, signal splits and link capacity increase variables. In order to ensure the delays and queues are acceptable at equilibrium condition, the link flows that approach to the signalized intersection must satisfy capacity constraints, i.e., capacity constraints are given as follows:

$$v_a(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\gamma}) \leq p_a s_a(\lambda_a, \gamma_a), a \in \bar{A}. \quad (8)$$

Furthermore, the green time at signal-controlled intersections and O-D demands should satisfy some linear constraints, given as follows:

$$\lambda_{\min} \leq \lambda_a \leq \lambda_{\max}, a \in \bar{A}; \quad (9)$$

$$\mu_{rs} \geq \mu_0, \forall r, s, \quad (10)$$

where: λ_{\min} is the minimum green split; λ_{\max} is the maximum green split; μ_0 is the minimum O-D demand multiplier.

For simplicity, lost time is not considered in the paper. Thus green splits must satisfy the following relationship:

$$\sum_{n=1}^{N_j} \lambda_{nj} = 1, j \in J, \quad (11)$$

where: N_j denotes the preset number of phases on signalized intersection j .

Ge et al. (2003) formulated a bi-level program to model the network capacity with SUE problem. But, in his model, the O-D multiplier associate with each O-D demand is assumed to increase at the same rate, which may restrict the maximum throughput of reserve capacity. According to the definition of reserve capacity proposed by Gao and Song (2002), the O-D demand multipliers between each O-D pair could be different. This concept relaxes the common multiplier requirement in the reserve capacity concept (Wong, Yang 1997) by allowing the maximum throughput to be scaled by individual O-D pairs (Chen, Kasikitwiwat 2011). Following their work, the mathematical programming for reserve capacity with SUE problem is formulated as follows:

Model 1:

$$\max z = \sum_{\forall r,s} \mu_{rs} q_{rs}^0 \quad (12a)$$

$$\text{s.t.} \begin{cases} v_a(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{y}) \leq p_a s_a(\lambda_a, y_a), a \in \bar{A}; \\ \mu_{rs} \geq \mu_0, \forall r, s; \\ \sum_{n=1}^{N_j} \lambda_{nj} = 1, j \in J; \\ \lambda_{\min} \leq \lambda_a \leq \lambda_{\max}, a \in \bar{A}; \\ y_a \geq 0, \forall a \in A, \end{cases} \quad (12b)$$

where: $v_a(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{y})$ is obtained by solving the following logit-based SUE problem:

$$\min z = \sum_{a \in A} v_a t_a(v_a, \lambda_a, y_a) - \sum_a \int_0^{v_a} t_a(x, \lambda_a, y_a) dx - \sum_{r,s} \mu_{rs} q_{rs}^0 S_{rs}(\mathbf{c}^{rs}(\mathbf{v})). \quad (13)$$

Model 2:

In order to accommodate the gradually growing travel demand, measures such as network design are usually taken by the network authorities to improve network performance. Classically this kind of design problems is considered in three forms: CND problem which deals with the continuous capacity expansion of the existing street; DND problem which deals with adding new streets or lanes to the existing streets and Mix Network Design (MND) problem which deals with both discrete and CND variables (Miandoabchi, Farahani 2011). A brief review of the three network design problems and its solution algorithms could be found in Table 1. Since in CND problem, all designed variables are continuous in their feasible regions, it is thus could be solved with gradient-based solution algorithm where multiple methods such as SAB method, iterative optimization assignment, projected quasi-Newton method and augmented Lagrangian algorithm are developed previously as can be seen in Table 1. The designed parameters in the DND and MND problem, however, are discrete (integers) instead of continuous, which as a re-

Table 1. Previous algorithms for solving network design problems

Problem	Name of the algorithm	Sources
CND problem	<i>Gradient-based algorithms</i>	
	SAB method	Connors et al. 2007; Sumalee 2007; Yang, Yagar 1994; Gao, Song 2002
	Iterative optimization assignment	Allsop, Charlesworth 1977
	projected quasi-Newton method	Chiou 2008
	Augmented Lagrangian algorithm	Meng et al. 2001
	<i>Derivative-free algorithms</i>	
	Viable global optimization method	Li et al. 2012
	Equilibrium decomposed optimization	Suwansirikul et al. 1987
	Genetic algorithm	Yin 2000
	Path based mixed-integer linear program	Wang, Lo 2010
DND problem	<i>Intelligent algorithms</i>	
	Ant system method	Poorzahedy, Abulghasemi 2005
	Genetic algorithm	Wu et al. 2012; Jeon et al. 2006; Drezner, Wesolowsky 2003
	Simulated annealing algorithm	Miandoabchi, Farahani 2011; Drezner, Wesolowsky 1997
	Particle swarm optimization	Zhang, Gao 2007
	<i>Non-intelligent algorithms</i>	
	SO-relaxation and UE principle	Wang et al. 2013
Support function concept	Gao et al. 2005	
MND problem	Genetic algorithm	Cantarella, Vitetta 2006; Wang, Deng 2015
	Cutting constraint method	Luathep et al. 2011

sult, makes the target function non-derivable and thus could not be solved with gradient-based method. Therefore, considering of the applicability of SAB method, this paper we only studied the network capacity with CND problem, formulated as:

$$\max z = \sum_{\forall r,s} \mu_{rs} q_{rs}^0 \quad (14a)$$

$$\text{s.t.} \begin{cases} \sum_{a \in A} G_a(y_a) \leq TI; \\ v_a(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{y}) \leq p_a s_a(\lambda_a, y_a) \quad a \in \bar{A}; \\ \sum_{n=1}^{N_j} \lambda_{nj} = 1, \quad j \in J; \\ \mu_{rs} \geq \mu_0, \quad \forall r, s; \\ \lambda_{\min} \leq \lambda_a \leq \lambda_{\max}, \quad a \in \bar{A}; \\ y_a \geq 0, \quad \forall a \in A, \end{cases} \quad (14b)$$

where: $v_a(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{y})$ is obtained by solving the network equilibrium problem (13).

3. Solution Algorithms for Solving Model 1 and Model 2

3.1. SAB Algorithm

3.1.1. General Description of SAB Algorithm

Similar to other forms of bi-level mathematical programming problems, our proposed bi-level reserve capacity problems (*Model 1*) and (*Model 2*), are also intrinsically nonconvex, and hence is very difficult to solve for a globally optimal solution. Besides, due to the implicit nonlinear functional term $v_a(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{y})$ in the up-level program, it is also troublesome to evaluate the changes in equilibrium link flows. To overcome this problem, Friesz *et al.* (1990) proposed the SAB method for bi-level problem in which a linear function is used to approximate the nonlinear function of equilibrium link flows. This method can be equally applied for solving our bi-level reserve capacity problem.

The concept of sensitivity analysis could be firstly traced back to the research by Hall (1978), Dafermos and Nagurney (1984). They investigated the direction of change when perturbations are added to the inputs of a UE traffic assignment model. Tobin and Friesz (1988) overcame the problem of non-uniqueness of the UE path flows by introducing an equivalent restricted program that has the desired uniqueness properties. Following their approach, Yang (1997) derived a gradient-based sensitivity analysis formula for the elastic demand (variable demand) network equilibrium problem. SAB method is previously used by Friesz *et al.* (1990) in solving bi-level network design problem. Subsequently, it was explored the efficiency and applicability for solving various bi-level optimization problems of transportation systems (see, e.g., Yang, Yagar 1994; Luo *et al.* 1995; Miyagi, Suzuki 1996; Bell, Iida 1997; Wang, Deng 2013).

The main idea of SAB method is to use a linear function to approximate the nonlinear function of equilibrium link flow $v_a(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{y})$ in the upper-level program (12b) and (14b). To do this, we must firstly obtain the derivatives of equilibrium link flows with respect

to perturbed parameters (i.e. O–D demand multipliers, signal splits, and link capacity increase variables). Assume the derivations have been calculated at current feasible point $(\mathbf{v}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*, \mathbf{y}^*)$, then, according to Taylor's formula, the implicit functional form $v_a(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{y})$ can be approximated as:

$$\begin{aligned} v_a(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{y}) \approx & v_a(\mathbf{u}^*, \boldsymbol{\lambda}^*, \mathbf{y}^*) + \\ & \sum_{a \in \bar{A}} \frac{\partial v_a(\mathbf{u}^*, \boldsymbol{\lambda}^*, \mathbf{y}^*)}{\partial \lambda_a} (\lambda_a - \lambda_a^*) + \\ & \sum_{a \in A} \frac{\partial v_a(\mathbf{u}^*, \boldsymbol{\lambda}^*, \mathbf{y}^*)}{\partial y_a} (y_a - y_a^*) + \\ & \sum_{\forall r,s} \frac{\partial v_a(\mathbf{u}^*, \boldsymbol{\lambda}^*, \mathbf{y}^*)}{\partial \mu_{rs}} (\mu_{rs} - \mu_{rs}^*). \end{aligned} \quad (15)$$

where: $v_a(\mathbf{u}^*, \boldsymbol{\lambda}^*, \mathbf{y}^*) = \mathbf{v}^*$.

Substituting equation (15) into the upper-level problem (12b) and (14b), the upper-level problems will become an ordinary linear or nonlinear programming problem with the variable link capacity increase, signal splits and O–D demand multipliers. This ordinary problems can be solved by the well-known simple method (for *Model 1*) or quadratic iterative method (for *Model 2*), thus one can get a new improved point $(\mathbf{v}^n, \boldsymbol{\mu}^n, \boldsymbol{\lambda}^n, \mathbf{y}^n)$ from which a new linear or nonlinear programming problem is again generated and again solved by the same method, and repeat the steps, finally, we can get the converge solution. The steps for implementing SAB method to solve the bi-level network reserve capacity problem are summarized as follows:

Step 1: Determine an initial set of the values $(\mathbf{u}^*, \boldsymbol{\lambda}^*, \mathbf{y}^*)$. Set $n = 0$.

Step 2: Using MSA method or Dial's algorithm to solve the lower-level SUE problem for given $\boldsymbol{\mu}^n, \boldsymbol{\lambda}^n, \mathbf{y}^n$ and hence get \mathbf{v}^n .

Step 3: Calculate the derivatives $\partial \mathbf{v}^n / \partial \boldsymbol{\mu}, \partial \mathbf{v}^n / \partial \boldsymbol{\lambda}$ and $\partial \mathbf{v}^n / \partial \mathbf{y}$ with sensitivity analysis method for SUE problem.

Step 4: Formulate local linear approximations of the upper-level link flow term $v_a(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{y})$ with the derivative information, and use simple method or quadratic iterative method to solve the resulted program to obtain the new O–D demands multipliers vector $\boldsymbol{\mu}^{n+1}$, signal splits vector $\boldsymbol{\lambda}^{n+1}$ and the link increase vector \mathbf{y}^{n+1} .

Step 5: If $\max \left| \frac{\boldsymbol{\mu}^{n+1} - \boldsymbol{\mu}^n}{\boldsymbol{\mu}^n} \right| \leq \varepsilon_1$, $\max \left| \frac{\boldsymbol{\lambda}^{n+1} - \boldsymbol{\lambda}^n}{\boldsymbol{\lambda}^n} \right| \leq \varepsilon_2$ and $\max \left| \frac{\mathbf{y}^{n+1} - \mathbf{y}^n}{\mathbf{y}^n} \right| \leq \varepsilon_3$, then stop, where $\varepsilon_1, \varepsilon_2$ and ε_3 are predetermined tolerance. Otherwise let $n = n + 1$ and return to Step 1.

The steps presented above imply that one difficulty in implementing the SAB method for our reserve capacity problem is to calculate the unknown terms $\partial \mathbf{v}^n / \partial \boldsymbol{\mu}, \partial \mathbf{v}^n / \partial \boldsymbol{\lambda}$ and $\partial \mathbf{v}^n / \partial \mathbf{y}$ (for notational simplicity, the superscript n will be neglected thereby). Generally, those derivatives can be obtained easily with sensitivity analysis method when the drivers make their route decisions based on Wardrop's first principle. However, noting that the low-level problem in program (12) and (14)

are logit-based SUE rather than DUE, the SAB method introduced by Wong and Yang (1997), Gao and Song (2002) can not be employed directly for our proposed reserve capacity. Ying and Miyagi (2001) formulated a computationally efficient link-based algorithm for sensitivity analysis of SUE by adopting Dial's algorithm (Dial 1971). The sensitivity analysis method proposed is very efficient in observing the changes of equilibrium link flows with respect to uncertainty parameters at logit-based SUE. This method can be further employed to calculate the derivatives of equilibrium link flows with respect to O-D demand multipliers, signal splits, and link capacity increase variables in equation (15). For completeness, the following we will explicitly present how to obtain these derivatives by using the sensitivity analysis method proposed by Ying and Miyagi (2001).

3.1.2. Method for Calculating $\partial \mathbf{v} / \partial \boldsymbol{\lambda}$, $\partial \mathbf{v} / \partial \mathbf{y}$

Sheffi (1985) proved that the SUE equilibrium solution is achieved if and only if $\mathbf{v} = (v_a)_{a \in A}$ can minimize equation (13). Namely, the equilibrium link solutions will be the extreme point of equation (13). Then we have:

$$\frac{\partial Z}{\partial v_a} = \left(-\sum_{r,s} \sum_k q_{rs} P_k^{rs} \delta_{a,k}^{rs} + v_a \right) \frac{dt_a}{dv_a} = 0. \quad (16)$$

In this paper, link performance functions developed by US Bureau of Public Roads (BPR, <http://www.fhwa.dot.gov/about>) are used, which is a strictly increasing function with link flows (i.e., $dt_a/dv_a > 0$). Therefore, the equation (16) can vanish only if:

$$v_a - \sum_{r,s} \sum_{k \in R_{rs}} \mu_{rs} q_{rs}^0 P_k^{rs} \delta_{a,k}^{rs} = 0. \quad (17)$$

Expand equation (17) we have:

$$v_a - \sum_{r,s} \mu_{rs} q_{rs}^0 \frac{\sum_{k \in R_{rs}} \exp(-\theta c_k^{rs}) \delta_{a,k}^{rs}}{\sum_{p \in R_{rs}} \exp(-\theta c_p^{rs})} = 0$$

$$v_a - \sum_{r,s} \mu_{rs} q_{rs}^0 \sum_{k \in R_{rs}} \frac{\partial S_{rs}}{\partial t_a} = 0. \quad (18)$$

Noting that left side of equation (18) is a function of signal splits vector $\boldsymbol{\lambda}$, O-D demand multiplier vector $\boldsymbol{\mu}$ and link capacity increase vector \mathbf{y} , whose explicit variables are included in the term $\sum_{rs} \mu_{rs} q_{rs}^0 \sum_{k \in R_{rs}} \partial S_{rs} / \partial t_a$. Then formula (18) can be rewritten as the following nonlinear equation:

$$F_a(\mathbf{v}, \mathbf{u}, \boldsymbol{\lambda}, \mathbf{y}) = v_a - \sum_{r,s} \mu_{rs} q_{rs}^0 \sum_{k \in R_{rs}} \partial S_{rs} / \partial t_a = 0, \quad a \in A. \quad (19)$$

Combining the nonlinear equations for other links, equation (19) can be rewritten in a vector form:

$$F(\mathbf{v}, \mathbf{u}, \boldsymbol{\lambda}, \mathbf{y}) = 0. \quad (20)$$

Differentiating both sides of equation (20) with respect to signal split vector $\boldsymbol{\lambda}$, yields:

$$\frac{\partial \mathbf{v}}{\partial \boldsymbol{\lambda}} = -(\nabla_{\mathbf{v}} \mathbf{F})^{-1} \frac{\partial \mathbf{F}}{\partial \boldsymbol{\lambda}}, \quad (21)$$

where:

$$\nabla_{\mathbf{v}} \mathbf{F} = \left(\frac{\partial F_a}{\partial v_b} \right)_{a,b} = \left(\delta_{a,b} - \sum_{r,s} \mu_{rs} q_{rs}^0 \frac{\partial (\partial S_{rs} / \partial t_a)}{\partial v_b} \right)_{a,b}, \quad (22)$$

where: $\delta_{a,b} = 1$ if $a = b$ and 0 otherwise.

Assuming link travel time is only influenced by its traffic flows, signal splits and link capacity increase variables, the interference between different links flows are not considered. Equation (22) can then be expended as:

$$\nabla_{\mathbf{v}} \mathbf{F} = \mathbf{E} - \left(\sum_{r,s} \mu_{rs} q_{rs}^0 \frac{\partial S_{rs}}{\partial t_a} \frac{\partial t_b}{\partial v_b} \right)_{a,b} = \mathbf{E} - \left(\sum_{r,s} \mu_{rs} q_{rs}^0 \left(\frac{-\theta \sum_{k \in R_{rs}} \exp(-\theta c_k^{rs}) \delta_{a,k}^{rs} \delta_{b,k}^{rs}}{\sum_{p \in R_{rs}} \exp(-\theta c_p^{rs})} + \frac{\theta \left(\sum_{k \in R_{rs}} \exp(-\theta c_k^{rs}) \delta_{a,k}^{rs} \right) \left(\sum_{k \in R_{rs}} \exp(-\theta c_k^{rs}) \delta_{b,k}^{rs} \right)}{\left(\sum_{p \in R_{rs}} \exp(-\theta c_p^{rs}) \right)^2} \right) \frac{\partial t_b}{\partial v_b} \right)_{a,b}, \quad (23)$$

where: \mathbf{E} is a unit matrix. Ying and Miyagi (2001) denoted that:

$$\frac{\sum_{k \in R_{rs}} \exp(-\theta c_k^{rs}) \delta_{a,k}^{rs}}{\sum_{p \in R_{rs}} \exp(-\theta c_p^{rs})} = \frac{v_a^{rs}}{\mu_{rs} q_{rs}^0}; \quad (24)$$

$$\frac{\sum_{k \in R_{rs}} \exp(-\theta c_k^{rs}) \delta_{a,k}^{rs} \delta_{b,k}^{rs}}{\sum_{p \in R_{rs}} \exp(-\theta c_p^{rs})} = \frac{v_{a,b}^{rs}}{\mu_{rs} q_{rs}^0}, \quad (25)$$

where: v_b^{rs} is the number of travelers from origin r to destination s who use link b ; $v_{a,b}^{rs}$ is the number of travelers from r to s who choose some path which contains both link a and link b . The link flows v_b^{rs} and $v_{a,b}^{rs}$ can be easily obtained by repeatedly operating Dial's algorithm. Then, we can obtain the value of $\nabla_{\mathbf{v}} \mathbf{F}$.

The term $\partial \mathbf{F} / \partial \boldsymbol{\lambda}$ in equation (21) can be calculated in the following way – let $(t_c)_{\lambda_m}$ be the derivatives with respect to λ_m when the link flows and link capacity increase are fixed. Then we have:

$$\frac{\partial \mathbf{F}}{\partial \boldsymbol{\lambda}} = \left(-\sum_{r,s} \mu_{rs} q_{rs}^0 \sum_{c \in A_m} \frac{\partial^2 S_{rs}}{\partial t_a \partial t_c} (t_c)_{\lambda_m} \right)_{a, \lambda_m}, \quad (26)$$

where: A_m is set of links affected by λ_m ; $(t_c)_{\lambda_m}$ can be obtained from the corresponding link cost functions. Note that the term $\partial^2 S_{rs} / \partial t_a \partial t_c$ has already been calculated in equation (23), the term $\partial \mathbf{F} / \partial \boldsymbol{\lambda}$ is thus computable. Following equation (22) and equation (26), the derivatives of equilibrium link flows with respect to signal splits can be obtained easily according to equation (21).

The derivative $\partial \mathbf{v} / \partial \mathbf{y}$ can be calculated in the similar way as it is for $\partial \mathbf{v} / \partial \boldsymbol{\lambda}$ by following the procedure (21–26).

3.1.3. Method for Calculating $\partial \mathbf{v} / \partial \boldsymbol{\mu}$

Differentiating both side of equation (20) with respect to O–D multiplier vector $\boldsymbol{\mu}$, we have:

$$\frac{\partial \mathbf{v}}{\partial \boldsymbol{\mu}} = -(\nabla_{\mathbf{v}} \mathbf{F})^{-1} \frac{\partial \mathbf{F}}{\partial \boldsymbol{\mu}}. \quad (27)$$

Noting that the $\nabla_{\mathbf{v}} \mathbf{F}$ has already been obtained in calculating $\partial \mathbf{v} / \partial \boldsymbol{\lambda}$ and $\partial \mathbf{v} / \partial \mathbf{y}$, the only unknown term in equation (27) is $\partial \mathbf{F} / \partial \boldsymbol{\mu}$. From equation (19) and (20), we have:

$$\frac{\partial \mathbf{F}}{\partial \boldsymbol{\mu}} = \left(-\sum_{r,s} q_{rs}^0 \frac{\sum_{k \in R_{rs}} \exp(-\theta c_k^{rs}) \delta_{a,k}^{rs}}{\sum_{p \in R_{rs}} \exp(-\theta c_p^{rs})} \right)_{r,s}. \quad (28)$$

The unknown term in right side of equation (28) is exactly the same as it is in equation (24), which can be computed by operating Dial’s algorithm (Dial 1971). Then, we can obtain the derivatives of link flows with respect to O–D demand multiplies based on equation (27).

3.2. GA for Solving Bi-Level Reserve Capacity Problem

SAB method is a gradient-based algorithm that solving the bi-level problems by finding the directions that minimize the target function with gradient information obtained. Although the gradient-based algorithm is very efficient in obtaining a satisfactory solution, it easily get into local optimality, that is, the result given by the sensitivity method may not be the global optimal one (Ge *et al.* 2003; Zhang, Ge 2004). In order to test the efficiency and accuracy of the SAB method, the GA, which is another method commonly applied for the bi-level network capacity is presented in this paper to make a comparison. GA is a random search techniques based on Darwinian evolution that is originally introduced by Holland (1992). It searches the optimal solution in the feasible region through natural principles and selection. GA is best known for its simple form, less restriction and more powerful search for improvement. It has been applied to a wide range of research area including engineering, sciences, and commerce (Mathew, Sharma 2009). Generally, GA can find the approximate global optimal solution if the parameters are well controlled and enough generation is guaranteed. It has already been widely used to solve various bi-level programs in transportation research (Ge *et al.* 2003; Miandoabchi, Farahani 2011; Chen *et al.* 2006; Yin 2000; Chootinan *et al.* 2005).

Since the typical procedures (reproduction, crossover and mutation) and discussions for GA are extensively reviewed and studied in numerous literatures, no additional explanations will be provided for them in this paper. In this study the chromosomes are represented as a string of real numbers with a length that is equal to the design parameters. In order to ensure the optimal link flow solutions do not violate the capacity constraints (8), a penalty term is added to the mapping functions to calculate the fitness of every individual in the population.

The fitness function is formulated as:

$$f(x) = \sum_{r,s} \mu_{rs} q_{rs}^0 - \beta \arg \max_{a \in A} (0, v_a - p_a s_a(\lambda_a, y_a)), \quad (29)$$

where: the term $\beta \arg \max_{a \in A} (0, v_a - p_a s_a(\lambda_a, y_a))$ is to penalize those individuals violating the capacity constraints; β is the penalty coefficient.

The GA-based procedure for bi-level problem can be described as follows:

Step 1: Initialization. Define the GA parameters such as crossover probability, population size, mutation rate and the range of parameters. Randomly generate an initial population coded by real numerical strings.

Step 2: Using MSA method or Dial’s algorithm for SUE assignment with the potential solutions given by the population. Then evaluate the fitness of each chromosome in the population with equation (29).

Step 3: Perform GA operators (i.e. reproduction, crossover, and mutation) to create offspring; increase generation counter.

Step 4: Convergence test: check the stopping criterion. If the termination criterion is met, accept the best individual in population as the approximate satisfactory solution and stop; otherwise, return to Step 2.

4. A Numerical Example

In this part, we will demonstrate the general use of the proposed methods in an example network with numerical analysis. The example network shown in Fig. 1 is taken from Wong and Yang (1997), Gao and Song (2002). It has two O–D pairs, seven links and six nodes, of which nodes *E* and *F* are signal-controlled intersections. There are three paths, AEB and AFB and AEFB, for the O–D pair A–B while there is only one path, CEFD for O–D pair C–D. The current O–D demand from nodes A to B is 18 veh/min, and from nodes C to D is 6 veh/min. The input data taken from Gao and Song (2002) is summarized in Table 2. Suppose intersections *E* and *F* are controlled by two independent splits, λ_1 and λ_2 .

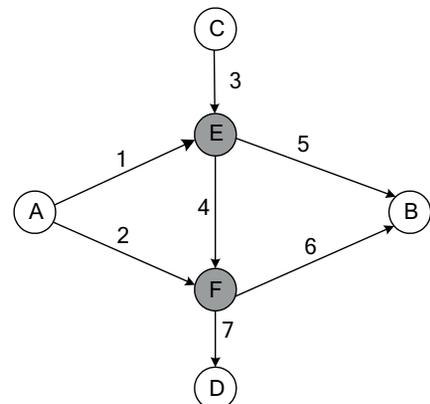


Fig. 1. The example road network (Wong, Yang 1997; Gao, Song 2002)

Table 2. Input data to the example network

Link number a	1	2	3	4	5	6	7
Free-flow time t_a^0	2.0	1.0	2.0	3.0	1.0	2.0	1.0
Saturation flow k_a	24	30	30	35	24	30	30
Parameter d_a	2	2	2	2	2	2	2
Link travel cost	$t_a(y_a, \lambda_a, y_a) = t_a^0 \left(1.0 + 0.5 \left(y_a / \lambda_a s_a \right)^2 \right)$						
Link capacity	$s_a(y_a, \lambda_a) = \lambda_a (k_a + y_a)$						
Investment function	$G_a(y_a) = 1.5 d_a (y_a)^2$						

Signal splits for link 1, 2, 3 and 4 are represented by $\lambda_1, \lambda_2, \lambda_3, \lambda_4$. The lower and upper bounds of signal splits are 0.05 and 0.95 respectively. The minimum O-D demand for each O-D pair is set as its current values, i.e., 18 veh/min for O-D pair (A, B) and 6 veh/min for O-D pair C-D. Therefore, all minimum O-D multipliers are fixed at 1. The maximum degree of saturation for all signal-controlled links is taken the same value of $p = 0.9$.

4.1. Example 1

Firstly, SAB method is used to solve the bi-level problem. Converge criteria are taken as 0.001 for all designed variables. The initial value for designed variables are set as $\mu_{AB} = \mu_{CD} = 1$ and $\lambda_1 = \lambda_2 = 0.5$ respectively. Table 3 lists the numerical results of Model 1 calculated with SAB method at $\theta = 0.5$. The convergence of SAB method is achieved in four iterations, very efficient in obtaining a satisfactory solution. In order to test whether sensitivity analysis method is trapped in local optimal solution, GA is employed to solve the reserve capacity problem to make a comparison. For GA procedure, the parameters are set as:

- real code GA is used and considered up to three decimal precision;
- population size 80;
- crossover operator is arithmetic crossover with a probability of 0.7;
- mutation rate, 0.25;
- penalty coefficient are fixed and set as $\beta = 20$;
- the maximal number of generation is 250.

Fig. 2 demonstrates that the optimal results calculated by GA are identical with SAB method. Therefore, both SAB method and GA can give the global best result for this small numeric example network. However, despite the fact that GA can help to find optimal solution, it converges very slowly and takes hundreds of generations to obtain the accurate optimal result. This is because the GA is a random search technique that optimizing the solutions based on nature selection. As is denoted previously, there is no efficient method for SUE problem and generally MSA method or Dial's algorithm is used which is very difficult to converge. Therefore, we use the number of SUE problems solved in finding the satisfactory solution as the performance indicator to compare the efficiency of GA and SAB method. From

Table 3. Numerical results with the SAB method

Iteration	λ_1	λ_2	μ_{AB}	μ_{CD}
1	0.500	0.500	1.000	1.000
2	0.778	0.759	1.937	1.000
3	0.778	0.776	1.950	1.000
4	0.778	0.776	1.950	1.000

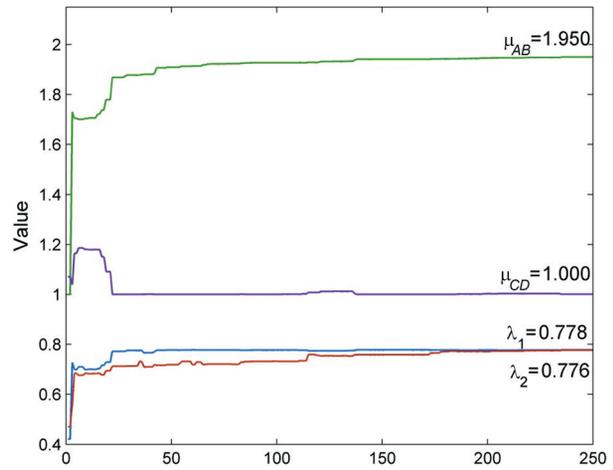
Fig. 2. Numerical results with GA at $\theta = 0.5$

Fig. 2, the number of SUE problems solved in GA can be easily calculated, which is $80 \times 250 = 20000$. The SAB method, however, only takes 4 iterations and 4 SUE problems are solved in obtaining the satisfactory solution. Therefore, compared with GA, SAB method reduces the computation complexity significantly, thus is much more preferable. In view of the efficiency of SAB method, it is employed for solving reserve capacity problem with different other θ . The corresponding numerical results are presented in Table 4.

The dispersion parameter θ in logit-based SUE represents how precisely a driver can correctly choose the shortest routes, the larger the θ , the higher the probability that a driver chooses the shortest routes. When θ approaches to infinite, the drivers will have perfect knowledge of the traffic and the SUE problem will turn into DUE problem. Table 4 demonstrates that when θ approaches to infinite (DUE) the total demand flow is

Table 4. Numerical results for Model 1 with different value of θ

θ	0.1	0.3	0.5	1	1.113	2	2.208	5	10	20	∞
λ_1	0.778	0.778	0.778	0.778	0.778	0.778	0.778	0.778	0.778	0.778	0.778
λ_2	0.614	0.722	0.776	0.807	0.808	0.810	0.810	0.810	0.810	0.81	0.810
λ_3	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222
λ_4	0.386	0.278	0.224	0.193	0.192	0.190	0.190	0.190	0.190	0.19	0.190
μ_{AB}	1.548	1.799	1.95	2.081	2.093	2.141	2.148	2.12	2.108	2.101	2.093
μ_{CD}	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1	1.000
Total demands	33.864	38.382	41.102	43.463	43.677	44.542	44.657	44.167	43.937	43.818	43.677
v_1	16.800	16.800	16.800	16.799	16.800	16.800	16.800	16.310	16.080	15.955	15.820
v_2	11.064	15.582	18.302	20.661	20.876	21.742	21.857	21.857	21.857	21.857	21.857
v_3	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
v_4	12.167	8.763	7.050	6.072	6.038	6.000	6.000	6.000	6.000	6.000	6.000
v_5	10.633	14.037	15.750	16.727	16.761	16.800	16.800	16.310	16.080	15.955	15.820
v_6	17.231	18.345	19.352	20.733	20.914	21.743	21.857	21.857	21.857	21.857	21.857
v_7	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000

43.677 veh/min and the corresponding optimal signal splits are $\lambda_1 = 0.778$, $\lambda_2 = 0.810$, $\lambda_3 = 0.222$, $\lambda_4 = 0.190$ respectively, the same results as is presented by Gao and Song (2002). However, out of our expectation, the numerical results in Table.4 demonstrate that the total demand is not the maximum when drivers make route choice based on DUE principle. As a matter of fact, it increases monotonously with θ and gets the best performance at $\theta = 2.208$, then decreases gradually as θ continue to increase. In other words, the increased quality of traveler information does not always contribute to the increase in network reserve capacity, and may even on the opposite, decreases the network capacity when more traffic information is provided. This is a paradox as we always struggle to find ways such as the Advanced Traveler Information System (ATIS) to improve the quality of travelers' information. Thus, if the character of reserve capacity with respect to traveler information quality isn't properly analyzed beforehand, it is very likely that the measures such as improving ATIS will reduce the network reserve capacity-making traffic congestion situation even worse.

Table 4 also indicates that the reserve capacity mainly perturbs with μ_{AB} , for μ_{CD} it is fixed with different θ . Therefore, by looking into the changes of route time and route flow between O–D pair A–B (see Fig. 3), we can take a deep insight into the fluctuating pattern of reserve capacity with different θ . According to logit-based discrete choice model, the difference of travel costs of used routes between each O–D pair varies inversely with the driver information quality. This can be seen from Fig. 3a. It shows that when driver information quality is low (θ is relatively small), there is a big deviation of travel cost between route AEB and route AFB. In order to achieve such a big difference of route travel time, the traffic flow on route AEB and route AFB is also small (Fig. 3b). As θ increases, the flow on route AEB and route AFB increase simultaneously to diminish the

difference of route travel time until maximum reserve capacity is reached at $\theta = 2.208$ (Fig. 3b). After that, the route AFB is saturated since it is the shortest route between O–D pair A–B while the flow on route AEB starts to decrease monotonously with respect to θ (Fig. 3b).

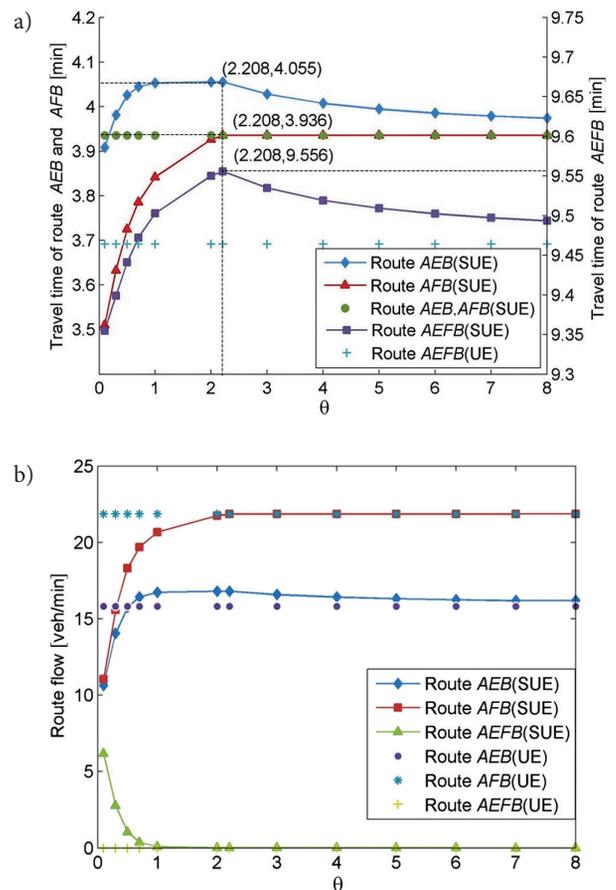


Fig. 3. Changes in network states: a – pattern of change in travel time; b – pattern of change in route flow

Meanwhile, the travel time of route AFB is getting closer and closer to that of route AEB and finally converge to the same value when θ approaches to infinite (i.e. DUE) (Fig. 3a). Consequently, we can draw a conclusion that the network reserve capacity is not always the maximum when drivers make route choice based on DUE principle because better information allows a large portion of demand to use the fast route, thus saturates the weakest link of that route, making it impossible to accommodate more traffic. Generally, traffic demand can be more evenly distributed among the routes between a O-D pair when traveler information quality is controlled at certain level, which allowing more demands to be served until flow on the key links reaches their capacity (Ge et al. 2003).

4.2. Example 2

Model 2 studies the reserve capacity with CND problem. Differing from the usual ones, it optimizes the allocation of investment for link capacity improvement to maximize the reserve capacity. SAB method is also applied to solve the bi-level problem considering of its efficiency. The converge criteria is taken the same as it is in example 1. Table 5 shows the numerical results when the total investment is set at 30. It denotes that for different θ , the total demands in Table 5 are all larger than they are in Table 4, a direct proof of the effectiveness of CND measure in improving the network reserve capacity.

In order to study the optimum pattern of link capacity increase, we calculate the congestion level of link 1 and link 2 at $TI = 0$, which is summarized on Table 6. The data in Table 5 and Table 6 demonstrate an interesting phenomenon. When driver information quality

is low ($\theta \leq 1$), link 1 is highly congested at maximum reserve capacity state at $TI = 0$ since $v_1/(0.9s_1) = 1$, while link 2 is uncongested (Table 6, the boxed digits are less than 1). The corresponding investment in this situation is mainly used to improve the capacity of congested link (link 1) while the capacity increase on parallel uncongested link (link 2) is very limited. As the level of driver information quality increases, say $\theta \geq 3$, link 2 becomes congested ($v_2/(0.9s_2) = 1$) while link 1 is no longer congested (Table 6, digits with dark background are less than 1). The investment at this time, however, is not only mainly used for congested link (link 2), but also a relative large proportion of investment is allocated for improving the capacity of uncongested link (e.g. link 1, digits with dark background in Table 5 shows), in particular when $\theta = 3$, the link capacity increase on uncongested link y_1 is nearly the same as it is on congested link y_2 with 2.142 and 2.221 respectively. The phenomenon is quite understandable: when quality of driver information is low, the drivers are not very sensitive to the improved network condition because of the limited information. As a result, the improved condition of one link or route will not significantly affects the drivers' decisions on choosing other links or routes. Consequently, to better increase the reserve capacity, the total investment in this case should be mainly used to improve the congested links. However, when the drivers have a good knowledge of the traffic condition, they will become highly sensitive to the improved network conditions. In this situation, if only the congested links get improved, they would attract large number of flows from other links, thus may cut down the overall effect of CND measure for improving network reserve capacity.

Table 5. Numerical results for Model 2 with total investment 30 (e.g. $TI = 30$)

θ	0.1	0.3	0.5	0.7	1	2	3 (max)	5	10	20	UE
λ_1	0.783	0.783	0.783	0.783	0.783	0.782	0.782	0.781	0.78	0.78	0.780
λ_2	0.588	0.711	0.772	0.797	0.808	0.812	0.812	0.812	0.813	0.813	0.813
μ_{AB}	1.751	2.017	2.173	2.25	2.299	2.327	2.330	2.327	2.318	2.311	2.304
μ_{CD}	1	1	1	1	1	1	1	1	1	1	1
Total demands	37.518	42.306	45.114	46.5	47.382	47.886	47.940	47.886	47.724	47.602	47.474
y_1	3.071	3.057	3.028	2.996	2.901	2.424	2.142	1.741	1.313	1.331	1.318
y_2	0.052	0.206	0.342	0.436	0.845	1.887	2.221	2.541	2.764	2.754	2.763
y_3	0.749	0.745	0.737	0.729	0.705	0.584	0.513	0.414	0.309	0.314	0.305
y_4	0.031	0.072	0.087	0.096	0.176	0.395	0.467	0.537	0.587	0.584	0.59
y_5	0	0	0	0	0	0	0	0.223	0.444	0.45	0.435
y_6	0.053	0.224	0.402	0.541	0.584	0.256	0.002	0	0	0	0
y_7	0	0	0	0	0	0	0	0	0	0	0

Table 6. The congestion level on link 1 and link 2 at maximum reserve capacity with no investment ($TI = 0$)

θ	0.1	0.3	0.5	0.7	1	2	3	5	10	20	UE
$v_1/(0.9s_1)$	1.000	1.000	1.000	1.000	1.000	1.000	0.987	0.971	0.957	0.949	0.941
$v_2/(0.9s_2)$	0.667	0.799	0.874	0.948	0.957	0.994	1.000	1.000	1.000	1.000	1.000

Therefore, when driver information quality is high, the investment should not only be used to improve the capacity of congested links, but also a relative large investment should be allocated for parallel uncongested links in order to achieve the maximum increase of network reserve capacity.

Fig. 4 describes the maximum total demands regarding to different total investment. It denotes that the total demand increases monotonously with total investment. Besides, the difference of total demand between SUE and UE reduces gradually when total investment increases. To make it more clearly, Fig. 5 depicts the difference of total demands between SUE and DUE with different investment when $\theta \geq 1$. The numerical results presented in Fig. 5 denote that the maximum difference of total demands between SUE and DUE varies inversely with respect to total investment. For example, the maximum difference of total demands between SUE and DUE results is 0.692 at $TI = 10$, less than it is at $TI = 70$ by 0.524, which is 0.168 correspondingly. Therefore, we can draw a conclusion from Figs 4–5 that CND can not only increase network reserve capacity effectively, but also helps to reduce the difference of reserve capacity between the assumptions of SUE and DUE. The more total investment is allocated, the less difference of maximum reserve between SUE and DUE is supposed to be.

The phenomenon can be explained by inspecting Table 7. As is mentioned previously, it is the perturbed demands between O–D pair A–B that mainly account for the change of reserve capacity, for O–D demands between C–D, it is always fixed at initial value. Consequently, according to the network topology, the total demands of the road network are account for by traffic flow on link 1 and link 2. In Table 7, $p_1\lambda_1s_1$ and $p_2\lambda_2s_2$ represent the physical capacity of link 1 and link 2 respectively. $p_1\lambda_1s_1 - v_1$ and $p_2\lambda_2s_2 - v_2$ denote the unused capacity of the link 1 and link 2 at DUE state. This definition implies that the more unused capacity on link 1 and link 2, the less efficient use of network physical capacity. Table 7 demonstrates that the unused capacity of link 2 is always approaching to zero with different total investment while the unused capacity in link 1 decreases dramatically as total investment increases. Consequently, we can draw a conclusion from Fig. 4 and Table 7 that the CND measure can not only help to improve the link capacity physically, but also serves to make more use of potential reserve capacity at DUE, thus reduces the difference of maximum reserve capacity between SUE and DUE.

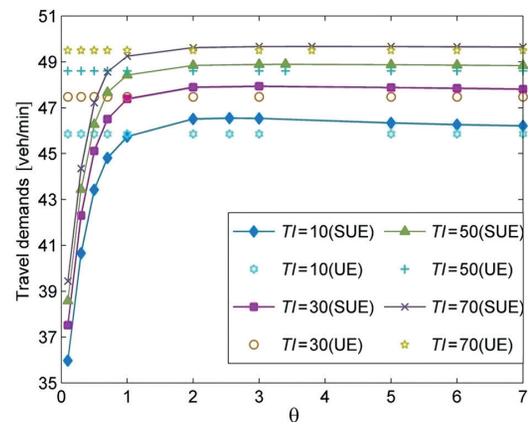


Fig. 4. Total demands at different θ with different total investment

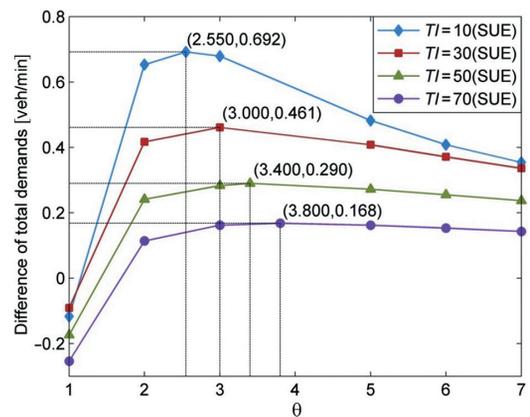


Fig. 5. Difference of total demands between SUE and corresponding DUE with different total investment

Conclusions

In this paper, two bi-level program models are formulated to model the reserve capacity problem with stochastic user route choice. One is reserve capacity with SUE problem, and the other combines reserve capacity with the CND problem. Two solution algorithms, i.e. the SAB and GA, are proposed to solve the bi-level problem network capacity problem and numerical results are compared with each other. The findings from the example network are:

1. Although GA can find a satisfactory solution for the network capacity problem, it is extremely time-consuming and computationally expensive due to the hard convergence of Dial’s algorithm method. The

Table 7. Optimal results of link 1 and link 2 for Model 2 at DUE (θ approach to infinite) with different total investment

TI	λ_1	λ_2	y_1	y_2	v_1	v_2	$p_1\lambda_1s_1$	$p_2\lambda_2s_2$	$p_1\lambda_1s_1 - v_1$	$p_2\lambda_2s_2 - v_2$
0	0.7778	0.8095	0	0	15.8226	21.8571	16.8005	21.8565	0.9779	-0.0006
10	0.7791	0.8113	0.7612	1.5981	16.782	23.0726	17.3623	23.0720	0.5803	-0.0006
30	0.78	0.8127	1.3178	2.7628	17.5102	23.9635	17.7731	23.9637	0.2629	0.0002
50	0.7807	0.8136	1.7009	3.562	18.0255	24.5762	18.0582	24.5754	0.0327	-0.0008
70	0.7819	0.8142	2.3760	4.0392	18.559	24.9437	18.5611	24.9432	0.0021	-0.0005

- proposed SAB method improves the efficiency of GA considerably by obtaining the decent directions that minimizing the target function, thus is much more preferable and practicable compared with GA.
2. Reserve capacity perturbs heavily with different quality of traveler's information. Generally, it reaches the maximum when driver information quality is controlled at certain level at SUE instead of at DUE. This is because, the drivers' perfect knowledge of traffic at DUE condition will leads to the early saturation of links on fast routes, preventing the further increase of travel demand due to capacity constraints, despite slower routes that still have unused capacity (Ge et al. 2003).
 3. When traveler information quality is low, the network investment is mainly used to improve the congested links in order to achieve the maximum increase of reserve capacity. However, when traveler information quality is high, only improve the capacity of congested link is not enough, a relative large proportion of the investment should also be allocated for 'parallel' uncongested links so as to maximize the reserve capacity.
 4. CND measure can not only help to maximize the reserve capacity physically, but also help to make more use of the potential road network capacity at DUE, thus reduces the difference in the reserve capacity between the assumptions of SUE and DUE. Therefore, city transportation decision-makers can combine the ATIS strategy and CND measure to improve the network reserve capacity to reduce the possible congestion.

Acknowledgements

This work was jointly supported by the National Key Technology Research and Development Program of China (Project No 2006BAJ18B03) and the Research and Innovation Project for College Graduate of Jiangsu Province (Project No CXZZ11_0165).

References

- Allsop, R. E.; Charlesworth, J. A. 1977. Traffic in a signal-controlled road network: an example of different signal timings including different routeing, *Traffic Engineering & Control* 18(5): 262–264.
- Bell, M. G. H.; Iida, Y. 1997. *Transportation Network Analysis*. Wiley. 226 p.
- Cantarella, G. E.; Vitetta, A. 2006. The multi-criteria road network design problem in an urban area, *Transportation* 33(6): 567–588.
<http://dx.doi.org/10.1007/s11116-006-7908-z>
- Ceylan, H.; Bell, M. G. H. 2004. Reserve capacity for a road network under optimized fixed time traffic signal control, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 8(2): 87–99.
<http://dx.doi.org/10.1080/15472450490437780>
- Chen, A.; Chootinan, P.; Wong, S. C. 2006. New reserve capacity model of signal-controlled road network, *Transportation Research Record* 1964: 35–41.
<http://dx.doi.org/10.3141/1964-05>
- Chen, A.; Kasikitwiwat, P. 2011. Modeling capacity flexibility of transportation networks, *Transportation Research Part A: Policy and Practice* 45(2): 105–117.
<http://dx.doi.org/10.1016/j.tra.2010.11.003>
- Chen, A.; Yang, H.; Lo, H. K.; Tang, W. H. 2002. Capacity reliability of a road network: an assessment methodology and numerical results, *Transportation Research Part B: Methodological* 36(3): 225–252.
[http://dx.doi.org/10.1016/S0191-2615\(00\)00048-5](http://dx.doi.org/10.1016/S0191-2615(00)00048-5)
- Chen, A.; Yang, H.; Lo, H. K.; Tang, W. H. 1999. A capacity related reliability for transportation networks, *Journal of Advanced Transportation* 33(2): 183–200.
<http://dx.doi.org/10.1002/atr.5670330207>
- Chiou, S.-W. 2008. A hybrid approach for optimal design of signalized road network, *Applied Mathematical Modelling* 32(2): 195–207.
<http://dx.doi.org/10.1016/j.apm.2006.11.007>
- Chiou, S.-W. 2007. Reserve capacity of signal-controlled road network, *Applied Mathematics and Computation* 190(2): 1602–1611. <http://dx.doi.org/10.1016/j.amc.2007.02.041>
- Chootinan, P.; Wong, S. C.; Chen, A. 2005. A reliability-based network design problem, *Journal of Advanced Transportation* 39(3): 247–270.
<http://dx.doi.org/10.1002/atr.5670390303>
- Connors, R. D.; Sumalee, A.; Watling, D. P. 2007. Sensitivity analysis of the variable demand probit stochastic user equilibrium with multiple user-classes, *Transportation Research Part B: Methodological* 41(6): 593–615.
<http://dx.doi.org/10.1016/j.trb.2006.11.003>
- Dafermos, S.; Nagurney, A. 1984. Sensitivity analysis for the asymmetric network equilibrium problem, *Mathematical Programming* 28(2): 174–184.
<http://dx.doi.org/10.1007/BF02612357>
- Damberg, O.; Lundgren, J. T.; Patriksson, M. 1996. An algorithm for the stochastic user equilibrium problem, *Transportation Research Part B: Methodological* 30(2): 115–131.
[http://dx.doi.org/10.1016/0191-2615\(95\)00026-7](http://dx.doi.org/10.1016/0191-2615(95)00026-7)
- Dial, R. B. 1971. A probabilistic multipath traffic assignment model which obviates path enumeration, *Transportation Research* 5(2): 83–111.
[http://dx.doi.org/10.1016/0041-1647\(71\)90012-8](http://dx.doi.org/10.1016/0041-1647(71)90012-8)
- Drezner, Z.; Wesolowsky, G. O. 2003. Network design: selection and design of links and facility location, *Transportation Research Part A: Policy and Practice* 37(3): 241–256.
[http://dx.doi.org/10.1016/S0965-8564\(02\)00014-9](http://dx.doi.org/10.1016/S0965-8564(02)00014-9)
- Drezner, Z.; Wesolowsky, G. O. 1997. Selecting an optimum configuration of one-way and two-way routes, *Transportation Science* 31(4): 386–394.
<http://dx.doi.org/10.1287/trsc.31.4.386>
- Fisk, C. 1980. Some developments in equilibrium traffic assignment, *Transportation Research Part B: Methodological* 14(3): 243–255. [http://dx.doi.org/10.1016/0191-2615\(80\)90004-1](http://dx.doi.org/10.1016/0191-2615(80)90004-1)
- Friesz, T. L.; Tobin, R. L.; Cho H.-J.; Mehta, N. J. 1990. Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints, *Mathematical Programming* 48(1–3): 265–284.
<http://dx.doi.org/10.1007/BF01582259>
- Gao, Z. Y.; Song, Y. F. 2002. A reserve capacity model of optimal signal control with user-equilibrium route choice, *Transportation Research Part B: Methodological* 36(4): 313–323. [http://dx.doi.org/10.1016/S0191-2615\(01\)00005-4](http://dx.doi.org/10.1016/S0191-2615(01)00005-4)
- Gao, Z.; Wu, J.; Sun, H. 2005. Solution algorithm for the bi-level discrete network design problem, *Transportation Research Part B: Methodological* 39(6): 479–495.
<http://dx.doi.org/10.1016/j.trb.2004.06.004>

- Ge, Y. E.; Zhang, H. M.; Lam, W. H. K. 2003. Network reserve capacity under influence of traveler information, *Journal of Transportation Engineering* 129(3): 262–270. [http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:3\(262\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(2003)129:3(262))
- Hall, M. A. 1978. Properties of the equilibrium state in transportation networks, *Transportation Science* 12(3): 208–216. <http://dx.doi.org/10.1287/trsc.12.3.208>
- Holland, J. H. 1992. *Adaptation in Natural and Artificial Systems*. A Bradford Book. 211 p.
- Jeon, K.; Lee, J. S.; Ukkusuri, S.; Waller, S. 2006. Selectorecombinative genetic algorithm to relax computational complexity of discrete network design problem, *Transportation Research Record* 1964: 91–103. <http://dx.doi.org/10.3141/1964-11>
- Kasikitwiwat, P.; Chen, A. 2005. Analysis of transportation network capacity related to different system capacity concepts, *Journal of the Eastern Asia Society for Transportation Studies* 6: 1439–1454.
- Li, C. M.; Yang, H.; Zhu, D. L.; Meng, Q. 2012. A global optimization method for continuous network design problems, *Transportation Research Part B: Methodological* 46(9): 1144–1158. <http://dx.doi.org/10.1016/j.trb.2012.05.003>
- Luathep, P.; Sumalee, A.; Lam, W. H. K.; Li, Z.-C.; Lo, H. K. 2011. Global optimization method for mixed transportation network design problem: a mixed-integer linear programming approach, *Transportation Research Part B: Methodological* 45(5): 808–827. <http://dx.doi.org/10.1016/j.trb.2011.02.002>
- Luo, Z.-Q.; Pang, J.-S.; Ralph, D. 1995. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press.
- Maher, M. J.; Hughes, P. C. 1997. A probit-based stochastic user equilibrium assignment model, *Transportation Research Part B: Methodological* 31(4): 341–355. [http://dx.doi.org/10.1016/S0191-2615\(96\)00028-8](http://dx.doi.org/10.1016/S0191-2615(96)00028-8)
- Mathew, T. V.; Sharma, S. 2009. Capacity expansion problem for large urban transportation networks, *Journal of Transportation Engineering* 135(7): 406–415. [http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(2009\)135:7\(406\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(2009)135:7(406))
- Meng, Q.; Yang, H.; Bell, M. G. H. 2001. An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem, *Transportation Research Part B: Methodological* 35(1): 83–105. [http://dx.doi.org/10.1016/S0191-2615\(00\)00016-3](http://dx.doi.org/10.1016/S0191-2615(00)00016-3)
- Miandoabchi, E.; Farahani, R. Z. 2011. Optimizing reserve capacity of urban road networks in a discrete network design problem, *Advances in Engineering Software* 42(12): 1041–1050. <http://dx.doi.org/10.1016/j.advengsoft.2011.07.005>
- Miyagi, T.; Suzuki, T. 1996. A Ramsey price equilibrium model for urban transit system: a bilevel programming approach with transportation network equilibrium constraints, in *World Transport Research: Proceedings of the 7th World Conference on Transport Research*, 16–21 July 1995, Sydney, Australia 2: 65–78.
- Poorzahedy, H.; Abulghasemi, F. 2005. Application of ant system to network design problem, *Transportation* 32(3): 251–273. <http://dx.doi.org/10.1007/s11116-004-8246-7>
- Powell, W. B.; Sheffi, Y. 1982. The convergence of equilibrium algorithms with predetermined step sizes, *Transportation Science* 16(1): 45–55. <http://dx.doi.org/10.1287/trsc.16.1.45>
- Prashker, J. N.; Bekhor, S. 2004. Route choice models used in the stochastic user equilibrium problem: a review, *Transport Reviews* 24(4): 437–461. <http://dx.doi.org/10.1080/0144164042000181707>
- Sheffi, Y. 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice Hall. 416 p.
- Sheffi, Y.; Powell, W. B. 1982. An algorithm for the equilibrium assignment problem with random link times, *Networks* 12(2): 191–207. <http://dx.doi.org/10.1002/net.3230120209>
- Sumalee, A. 2007. Multi-concentric optimal charging cordon design, *Transportmetrica* 3(1): 41–71. <http://dx.doi.org/10.1080/18128600708685667>
- Suwansirikul, C.; Friesz, T. L.; Tobin, R. L. 1987. Equilibrium decomposed optimization: a heuristic for the continuous equilibrium network design problem, *Transportation Science* 21(4): 254–263. <http://dx.doi.org/10.1287/trsc.21.4.254>
- Tobin, R. L.; Friesz, T. L. 1988. Sensitivity analysis for equilibrium network flow, *Transportation Science* 22(4): 242–250. <http://dx.doi.org/10.1287/trsc.22.4.242>
- Vincent, R. A.; Mitchell, A. I.; Robertson, D. I. 1980. *User Guide to TRANSYT Version 8*. TRRL Report LR888. Transport and Road Research Laboratory, Crowthorne, Berkshire England. 86 p.
- Wang, D. Z. W.; Lo, H. K. 2010. Global optimum of the linearized network design problem with equilibrium flows, *Transportation Research Part B: Methodological* 44(4): 482–492. <http://dx.doi.org/10.1016/j.trb.2009.10.003>
- Wang, J.; Deng, W. 2013. Reserve capacity on multiphase signalized road network with stochastic user equilibrium, in *TRB 92nd Annual Meeting Compendium of Papers*, 13–17 January 2013, Washington, DC, USA, 1–18. (DVD).
- Wang, J.; Deng, W. 2015. Optimizing capacity of signalized road network with reversible lanes, *Transport*. <http://dx.doi.org/10.3846/16484142.2014.994227> (in press).
- Wang, S.; Meng, Q.; Yang, H. 2013. Global optimization methods for the discrete network design problem, *Transportation Research Part B: Methodological* 50: 42–60. <http://dx.doi.org/10.1016/j.trb.2013.01.006>
- Webster, F. V.; Cobbe, B. M. 1966. *Traffic Signals*. Road Research technical paper No 56. London, England. 111 p.
- Wong, S. C.; Yang, H. 1997. Reserve capacity of a signal-controlled road network, *Transportation Research Part B: Methodological* 31(5): 397–402. [http://dx.doi.org/10.1016/S0191-2615\(97\)00002-7](http://dx.doi.org/10.1016/S0191-2615(97)00002-7)
- Wu, J.; Lu, H. P.; Yu, X. X.; Bian, C. Z. 2012. Genetic algorithm for multiuser discrete network design problem under demand uncertainty, *Mathematical Problems in Engineering* 2012: 1–17. <http://dx.doi.org/10.1155/2012/686272>
- Yang, H. 1997. Sensitivity analysis for the elastic-demand network equilibrium problem with applications, *Transportation Research Part B: Methodological* 31(1): 55–70. [http://dx.doi.org/10.1016/S0191-2615\(96\)00015-X](http://dx.doi.org/10.1016/S0191-2615(96)00015-X)
- Yang, H.; Bell, M. G. H.; Meng, Q. 2000. Modeling the capacity and level of service of urban transportation networks, *Transportation Research Part B: Methodological* 34(4): 255–275. [http://dx.doi.org/10.1016/S0191-2615\(99\)00024-7](http://dx.doi.org/10.1016/S0191-2615(99)00024-7)
- Yang, H.; Yagar, S. 1994. Traffic assignment and traffic control in general freeway-arterial corridor systems, *Transportation Research Part B: Methodological* 28(6): 463–486. [http://dx.doi.org/10.1016/0191-2615\(94\)90015-9](http://dx.doi.org/10.1016/0191-2615(94)90015-9)
- Yin, Y. F. 2000. Genetic-algorithms-based approach for bilevel programming models, *Journal of Transportation Engineering* 126(2): 115–120. [http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(2000\)126:2\(115\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(2000)126:2(115))

- Ying, J. Q.; Miyagi, T. 2001. Sensitivity analysis for stochastic user equilibrium network flows – a dual approach, *Transportation Science* 35(2): 124–133.
<http://dx.doi.org/10.1287/trsc.35.2.124.10137>
- Zhang, H.; Gao, Z. 2007. Two-way road network design problem with variable lanes, *Journal of Systems Science and Systems Engineering* 16(1): 50–61.
<http://dx.doi.org/10.1007/s11518-007-5034-x>
- Zhang, H. M.; Ge, Y. E. 2004. Modeling variable demand equilibrium under second-best road pricing, *Transportation Research Part B: Methodological* 38(8): 733–749.
<http://dx.doi.org/10.1016/j.trb.2003.12.001>
- Zhang, P.; Li, W.-Q.; Chang, Y.-L. 2010. Reserve capacity model for urban road network with variable lanes, *Journal of Southwest Jiaotong University* 45(2): 255–260 (in Chinese).