TIME-DEPENDENT TRANSPORT NETWORK DESIGN: A STUDY ON BUDGET SENSITIVITY

Hong K. LO
Associate Professor
Department of Civil Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, PRC
E-mail: cehklo@ust.hk

Wai Yuen SZETO
Postdoctoral Research Associate
Department of Civil Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, PRC
E-mail: ceszeto@ust.hk

Abstract: Planning transport network improvements often involve three parties: network users, private toll road operators, and the government. Each of these parties has distinctive objectives that are often in conflict with each other. This study develops a single-level optimization program to design the network improvements over time and to study the tradeoff among these parties for a range of budget values. Three numerical studies of a small network are provided to illustrate the equity issues among users and among private toll road operators, and the cost-effectiveness of the design from the government perspective.

Key Words: continuous network design problem, time-dependent user-equilibrium assignment

1. INTRODUCTION

Transport infrastructure is in an active phase of planning and development in many parts of Asia. For example, on a regional scale, we are pondering how and when should Hong Kong be better linked with the Pearl River Delta. Specifically, should the Hong Kong-Macau-Zhuhai bridge be built? As transport infrastructure projects are expensive, in terms of constrained government expenditures, they must be carefully scrutinized for cost-effectiveness. In addition to identifying the highway link to be added or improved, the timing and scaling of the improvements are important considerations. Traditionally, this analysis belongs to the discipline of transport network design. Yang and Bell (1998) provided a recent review on models and solution approaches on this discipline.

Past efforts (Friesz et al., 1993; Wong and Yang, 1997; Gartner and Stamatiadis, 2001; Maher et al., 2001; Meng et al., 2001; Lam et al., 2002) on transport network design focused on optimizing the network for a certain future time without clearly defining the time dimension of the planning horizon within the formulation. In fact, transport infrastructure projects are not a one-time event, but will be in operation far into the future. Meanwhile, travel demands will change over time, as will the network itself and the tolls. Furthermore, the costs and benefits of transport network improvements accrue over a long period of time. It is important to incorporate the time dimension into the network analysis, so that one can design for the optimal project initiation time, phasing, scaling, toll strategies, and financial arrangements over the planning horizon.

Recently, Lo and Szeto (2003) introduced the time dimension to one type of network design problems, namely the continuous network design problem (CNDP), and referred to this extension as the CNDP-T. This extension considers not only the time-dependent user-equilibrium assignment with elastic demands but also time-dependent travel demands, changing land-use patterns, and the gradually upgraded network during the planning horizon. In particular, this earlier study focused on comparing the CNDP-T formulation with the traditional static formulation, and highlighting the benefits of CNDP-T.

In this study, we extend the CNDP-T formulation to capture the effects of capacity enhancements and tolls on the increase in travel demand. This extended model also carries the same characteristics as the one proposed by Lo and Szeto (2003) and includes the CNDP formulation as a special case. The extended model is formulated as a single-level mathematical program, allowing the use of existing nonlinear programming techniques for solutions.

Planning transport network improvements often involve three parties: network users, private toll road operators, and the government. Each of these parties has distinctive objectives:

1. Users are concerned with their travel costs;
2. Private toll road operators are concerned with their revenues, and;
3. The government is concerned with the cost-effectiveness of the design.

With the proposed model, we illustrate the equity issues among users and among private toll road operators, and the cost-effectiveness of the design from the government perspective through three numerical examples. The results show that there is a tradeoff among these parties in the CNDP-T. Careful considerations must be included in time-dependent transport network design so as to strike a balance in benefit distribution among these parties.

The outline of this paper is as follows. Section 2 formulates the CNDP-T. Section 3 depicts the considerations of users, private toll road operators, and the government in the CNDP-T. Section 4 contains the numerical studies. Finally, Section 5 provides some concluding remarks.

2. FORMULATION

We consider a general transportation network with multiple Origin-Destination (OD) flows over the planning horizon $\left[0,T\right]$. The horizon is divided into N equal design intervals or design periods.

The following notations are adopted throughout this paper:

Set notations

RS	set of OD pairs
P^{rs}	set of paths for travelers connecting OD pair rs
A	set of links
B	set of toll links, $B \subset A$

Indices

rs	OD pair, $rs \in RS$
p	path between OD pair rs, $p \in P^{rs}$
a	link, $a \in A$
b	toll link, $b \in B$
τ	period, $\tau \in \left\{1,..,N\right\}$

Variables to be determined

$f_{p,\tau}^{rs}$	representative hourly flow on path p between OD pair rs in period τ

$y_{a,\tau}$	capacity enhancement in period τ
$\mathbf{f}$	column vector of $\left[f_{p,\tau}^{rs}\right]$
$\mathbf{y}$	column vector of $\left[y_{a,\tau}\right]$
$\mathbf{s}$	column vector of slack variables
$\mathbf{x}$	column vector of $\left[\mathbf{f}^T, \mathbf{y}^T, \mathbf{s}^T\right]^T$

<u>Parameters given</u>

α	parameter in link performance function
β	parameter in link performance function
k	construction cost parameter in construction cost function
n	factor converting the consumer surplus from an hourly basis to a period basis
ψ	cost of unit travel time.
B_τ	budget provided by the government at the beginning of period τ
h^{rs}	growth rate of potential demand between OD pair rs
c_a	initial capacity of link a
u_a	upper bound of the capacity of link a
t_a^0	free flow travel time of link a
δ_a^p	link-path incidence indicator, $\delta_a^p = 1$ if a is on p, $\delta_a^p = 0$ otherwise
$\tilde{q}_1^{rs}$	potential demand between OD pair rs in the first period
$\rho_{b,\tau}$	toll on link b in period τ

<u>Functions of capacity enhancements and equilibrium path flows</u>

$v_{a,\tau}$	representative hourly flow on link a in period τ
$t_{a,\tau}$	travel time on link a in period τ
q_τ^{rs}	representative hourly travel demand between OD pair rs in period τ
$\eta_{p,\tau}^{rs}$	path travel cost for travelers taking path p between OD pair rs in period τ
π_τ^{rs}	the minimum travel cost between OD pair rs in period τ
m^{rs}	function describing the effect of capacity enhancements on the travel demand function of OD pair rs
$D_\tau^{rs}(.)$	continuous travel demand function for OD pair rs in period τ
$\mathbf{h}(.)$	vector function

<u>Functions of capacity enhancements</u>

R_τ	unspent cumulative allocated budgets saved for the future period after τ
$g_a(.)$	improvement cost function of link a

<u>Function of potential demand</u>

$G^{rs}(.)$	potential demand function for OD pair rs

The following assumptions are made in this study:

1) Traffic assignment follows the user-equilibrium criterion.
2) The potential demand growth over time is known.
3) Only link widening is considered.
4) The link cost function is separable.
5) The interest, inflation, and benefit discount rates are taken as zero for simplicity.

6) Tolls are constant over time.
7) The budget in each period is known.

In particular, the first assumption is suitable for situations where (i) the length of each period is relatively long so that equilibrium can be obtained within each period and (ii) each traveler has perfect information about the network state as is typically assumed in deterministic traffic assignment. In planning, the time duration between each design period is usually in a multiple of five years, which is long enough for equilibrium to occur. The effect of the construction transition period, which is relatively short, is ignored. The above assumptions are not restrictive from a modeling perspective, but are adopted merely to simplify the analysis. They can be relaxed in future studies.

2.1 Objective Function

In this study, we adopt consumer surplus (CS) as the performance measure, which internalizes the effect of network congestion and the public's propensity to travel. Consumer surplus measures the difference between what consumers would be willing to pay for travel and what they actually pay. For the same network and demand characteristics, a higher consumer surplus implies a better performing system. Mathematically, it is expressed as:

$$TCS = \sum_{\tau} \sum_{rs} CS_{\tau}^{rs} , \tag{1}$$

$$CS_{\tau}^{rs} = n \left[\int_0^{q_{\tau}^{rs}} D_{\tau}^{rs-1}(v)\, dv - \pi_{\tau}^{rs} q_{\tau}^{rs} \right], \tag{2}$$

where TCS is the total consumer surplus; CS_{τ}^{rs} is the CS of OD pair rs in period τ; n is a factor converting the consumer surplus from an hourly basis to a period basis; q_{τ}^{rs} and π_{τ}^{rs} are, respectively, the travel demand and the minimum travel cost between OD pair rs in period τ. The first term in the square bracket in (2) is the total travel cost for the demand q_{τ}^{rs} that they would be willing to pay whereas the second term is the total travel cost that they actually pay. The total consumer surplus (TCS) in (1) is obtained by summing the consumer surplus measure over the planning horizon for all OD pairs, representing the overall network consumer surplus over the planning horizon.

2.2 Constraints
2.2.1 Time-dependent traffic assignment constraint

The CNDP-T includes N time-dependent traffic assignment sub-problems. For each sub-problem, travelers are assumed to follow the Wardrop's principle (1952). This principle requires that path p between OD pair rs will not be used if its travel cost is longer than the minimum travel cost between OD pair rs. Conversely, any used path p must have its travel cost equal to the minimum travel cost between OD pair rs. Mathematically, this principle for each sub-problem in period τ can be expressed as:

$$f_{p,\tau}^{rs} \left[\eta_{p,\tau}^{rs} - \pi_{\tau}^{rs} \right] = 0, \forall rs, p, \tau , \tag{3}$$

$$\eta_{p,\tau}^{rs} - \pi_{\tau}^{rs} \ge 0, \forall rs, p, \tau , \tag{4}$$

where $f_{p,\tau}^{rs}$ and $\eta_{p,\tau}^{rs}$ are respectively the representative hourly flow and the path travel cost for path p between OD pair rs in period τ. Equations (3) and (4) constitute the nonlinear complementarity conditions for the traffic assignment principle for each period τ. According to (3), if path p carries a positive flow in period τ, (i.e., $f_{p,\tau}^{rs} > 0$), then its associated path cost $\eta_{p,\tau}^{rs}$ must be equal to the lowest cost π_{τ}^{rs} through the condition $\left[\eta_{p,\tau}^{rs}-\pi_{\tau}^{rs}\right]=0$. Equation (4) ensures π_{τ}^{rs} to be the lowest cost among all the possible paths between OD pair rs in period τ.

Path travel costs and the minimum travel costs can be determined once the equilibrium path flows $\mathbf{f}$ and capacity enhancements $\mathbf{y}$ are known, expressed as:

$$v_{a,\tau}=\sum_{rs}\sum_{p} f_{p,\tau}^{rs}\delta_{a}^{p}, \forall a,\tau, \tag{5}$$

$$t_{a,\tau}=t_{a}^{0}\left[1+\alpha\left(\frac{v_{a,\tau}}{c_{a}+\sum_{i=1}^{\tau} y_{a,i}}\right)^{\beta}\right], \forall a,\tau, \tag{6}$$

$$\eta_{p,\tau}^{rs}=\sum_{a\in A/B}\psi t_{a,\tau}\delta_{a}^{p}+\sum_{b\in B}\left(\psi t_{b,\tau}+\rho_{b,\tau}\right)\cdot\delta_{b}^{p}, \forall rs,p,\tau, \tag{7}$$

where $v_{a,\tau}$ and $t_{a,\tau}$ are the representative hourly flow and the travel time on link a in period τ, respectively; $y_{a,\tau}$ is the capacity enhancement on link a in period τ, meaning that the capacity of link a is increased by $y_{a,\tau}$ units at the beginning of period τ; δ_{a}^{p} is a link-path incidence indicator, which equals one if link a is on path p, zero otherwise; t_{a}^{0} is the free flow travel time of link a; c_{a} is the capacity of link a before the capacity improvement; α,β are parameters of the link performance function; ψ is the cost of unit travel time.

Equation (5) states that link flow is obtained by summing the corresponding path flows on the link. Equation (6) is a typical link performance function. The summation term in (6) represents the total capacity enhancements of link a up to period τ. Therefore, the denominator inside the bracket denotes the link capacity in period τ after implementing the enhancements before and inclusive of period τ. When $\alpha=0.15$ and $\beta=4$, equation (6) is reduced to the typical Bureau of Public Roads (BPR) function. Equation (7) computes the path travel cost based on the corresponding link travel costs. The first term is the sum of the travel-time costs $\psi t_{a,\tau}$ on toll-free links on the path where as the second term is the sum of travel costs on toll links on the path in which the travel cost on each toll link is the sum of travel-time cost $\psi t_{b,\tau}$ and toll $\rho_{b,\tau}$ on that link.

Each traffic assignment sub-problem also includes the flow conservation and non-negativity conditions, expressed as:

$$\sum_{p} f_{p,\tau}^{rs}=q_{\tau}^{rs}, \forall rs,\tau, \tag{8}$$

$$f_{p,\tau}^{rs}\geq 0, \forall rs,p,\tau, \tag{9}$$

where q_{τ}^{rs} is the travel demand of OD pair rs in period τ.

The travel demand of OD pair rs in period τ, q_{τ}^{rs}, is not fixed but a function of the potential demand of that period $\tilde{q}_{\tau}^{rs}$ and its minimum travel costs π_{τ}^{rs} and $\pi_{\tau-1}^{rs}$:

$$q_{\tau}^{rs}=D_{\tau}^{rs}\left(\pi_{\tau}^{rs},\tilde{q}_{\tau}^{rs},m^{rs}\left(\pi_{\tau-1}^{rs},\pi_{\tau}^{rs}\right)\right), \forall rs,\tau, \tag{10}$$

where $D_\tau^{rs}(.)$ is the continuous travel demand function for OD pair rs in period τ ; $m^{rs}(.)$ is a function describing the effect of capacity enhancements on the functional form of the travel demand function of OD pair rs. Similar to Lo and Szeto (2003), the travel demand function of a period is generally decreasing with respect to the minimum travel cost of that period, implying that higher minimum travel costs lead to lower travel demands, and vice versa. Unlike Lo and Szeto (2003), however, the functional form of the demand function is changing over time to capture the increase in travel demand due to capacity enhancements. When the potential demand of each OD pair is fixed, the capacity enhancements on certain links lower the corresponding link times and OD travel costs, and hence attract the travel demands of corresponding OD pairs. The increase in travel demand depends on the ratio of the OD travel cost of the current period π_τ^{rs} to that of the previous period $\pi_{\tau-1}^{rs}$. The lower the ratio, the larger the increase in travel demand. In this study, the following travel demand function is adopted:

$$q_\tau^{rs} = \tilde{q}_\tau^{rs} - \gamma^{rs} \frac{\pi_\tau^{rs}}{\sqrt{\pi_{\tau-1}^{rs}}}, \tag{11}$$

where γ^{rs} is the parameter of the travel demand function of OD pair rs. This function is adopted for the purpose of illustration; other functional forms can be adopted in this framework without difficulty.

The potential demand per OD pair of each period represents the potential travel growth due to population growth and/or changes in land use patterns over time, expressed as:

$$\tilde{q}_\tau^{rs} = G^{rs}\left(\tilde{q}_{\tau-1}^{rs}\right), \tau > 1, \forall rs, \tag{12}$$

where $G^{rs}(.)$ is the potential demand function for OD pair rs. The potential demand in period τ is modeled to depend on the potential demand of period $\tau-1$ but not to depend on the traffic conditions. For simplicity, in this paper, the potential demand function is defined as:

$$\tilde{q}_\tau^{rs} = \tilde{q}_{\tau-1}^{rs}\left[1 + h^{rs}\right], \tag{13}$$

where h^{rs} is the growth rate of potential demand between OD pair rs.

This formulation shares some similarity with the Dynamic Traffic Assignment (DTA) problem, as both involve variable interactions across time in an intertwined manner. The difference is that in the CNDP-T, the equilibrium conditions hold at each discretized time in the planning horizon, typically several years apart; whereas in DTA, the equilibrium conditions hold for traffic at the same departure time, typically seconds or minutes apart (Lo and Szeto, 2002). In a sense, the CNDP-T is simpler. Traffic at one discretized time in the planning horizon does not interact with traffic at a different discretized time in the planning horizon, whereas in DTA, one must consider such interactions.

2.2.2 Link improvement constraints

For practical and physical reasons, roads or highways rarely have more then a few lanes. Therefore, the formulation includes link improvement constraints, expressed as:

$$c_a + \sum_\tau y_{a,\tau} \le u_a, \forall a, \tag{14}$$

$$y_{a,\tau} \ge 0, \forall a, \tau, \tag{15}$$

where u_a is the maximum allowable capacity of link a. The maximum allowable capacity constraint (14) limits the total capacity $c_a + \sum_{\tau} y_{a,\tau}$ of link a to be less than the maximum allowable capacity of that link u_a. Equation (15) is the non-negativity condition of capacity enhancements.

2.2.3 Budgetary constraints

The budgetary constraints can be stated as follows:

$$\sum_{a} g_a\left(y_{a,1}\right) + R_1 = B_1, \tag{16}$$

$$\sum_{a} g_a\left(y_{a,\tau}\right) + R_\tau = R_{\tau-1} + B_\tau, \forall \tau > 1, \tag{17}$$

where B_τ is the network improvement budget allocated by the government at the beginning of period τ; The variable R_τ represents the cumulative allocated funds that have not yet been spent in period τ and are available for use at the beginning of period $\tau+1$; The variable $g_a\left(y_{a,\tau}\right)$ is the cost of increasing the capacity of link a by $y_{a,\tau}$.

Equation (17) states that the sum of the total cost of improvements for period τ and the funds unspent or saved for the future years after τ is equal to the budget allocated by the government for period τ plus the funds saved from period 1 to $\tau-1$. Equation (16) describes the beginning of the planning horizon, assuming that no funds were saved before the planning horizon. This set of recursive equations permits the modeling of various budget and expenditure scenarios. For example, if the government provides a lump sum only at the beginning of the planning horizon, then $B_1 > 0$ and $B_\tau = 0,\ \tau > 1$. Then (16)-(17) can be reduced to $\sum_{\tau}\sum_{a} g_a\left(y_{a,\tau}\right) \le B_1$.

In this study, the construction cost function takes the following linear form:

$$g\left(y_{a,\tau}\right) = k t_a^0 y_{a,\tau}, \tag{18}$$

where k is a construction cost parameter. The improvement cost of a link is assumed to be proportional to the extent of the widening (and hence capacity gain) and its length (as represented by its free-flow travel time). Again, this function is adopted for illustration and simplicity; other functional forms can be adopted in this framework without difficulty. According to this equation, the higher the construction cost parameter, the higher the construction cost.

2.3 Formulating the CNDP-T as a single-level optimization program

The CNDP-T can be formulated as the following constrained nonlinear maximization program:

$$\max_{\mathbf{f},\mathbf{y}} TCS = n \sum_{\tau}\sum_{rs}\left[\int_0^{q_\tau^{rs}} D_\tau^{rs^{-1}}\left(v\right) dv - \pi_\tau^{rs} q_\tau^{rs}\right]$$

subject to

the time-dependent traffic assignment constraints (3)-(13);
the link improvement constraints (14)-(15), and;
the budgetary constraints (16)-(18).

The CNDP-T (1)-(18) is a direct extension of the CNDP-T in Lo and Szeto (2003), with the functional shape of the demand function considers the traffic conditions in the last planning period. As a result, they carry the same characteristics: i) The CNDP-T (1)-(17) is a generalization of the traditional CNDP and can be reduced to the traditional CNDP by setting $N = 1$ and removing the potential demand constraint, and ii) The CNDP-T (1)-(18) considers the time-dependent user-equilibrium assignment with elastic demands, time-dependent travel demands, changing land-use patterns, and the gradually upgraded network during the planning horizon. One can design for the optimal project initiation time, scale, and phasing over the planning horizon with this formulation.

Without loss of generality, the problem (1)-(18) can be written as follows:

$$\max \ TCS(\mathbf{x}) \tag{19}$$

subject to

$$\mathbf{h}(\mathbf{x}) = \mathbf{0}, \tag{20}$$

$$\mathbf{x} \geq \mathbf{0}, \tag{21}$$

where $\mathbf{x} = \left[\mathbf{f}^T, \mathbf{y}^T, \mathbf{s}^T\right]^T$; $\mathbf{h}(\mathbf{x})$ is a vector function of $\mathbf{x}$ representing the relationships (3)-(18); $\mathbf{s}$ is a column vector of slack variables. Slack variables $\mathbf{s}$ are introduced in the problem (19)-(21) to convert all inequality constraints in the problem (1)-(18) into equality constraints. The problem (19)-(21) constitutes a constrained nonlinear mathematical program that can be solved by existing optimization algorithms.

This study adopts the Generalized Reduced Gradient (GRG) algorithm (Abadie and Carpentier, 1969) to solve (19)-(21). The GRG method can work with mathematical programs with nonlinear objectives and nonlinear equality constraints. It combines the ideas of linearization and Wolfe's reduced gradient method. The latter is extended from the simplex method to solve nonlinear programming problems with linear equality constraints.

We note that the problem (19)-(21) contains non-convex constraints due to the time-dependent user-equilibrium conditions. Therefore, general-purpose nonlinear optimization codes will stop at local minima. How to take advantage of the special structure of this problem to establish stable, efficient solution algorithms is an important research topic. It is beyond the scope here and is left for future studies.

In summary, the proposed CNDP-T is to determine optimal capacity enhancements and the equilibrium flow patterns to maximize the consumer surplus over the planning horizon subject to the budgetary, link improvement, and time-dependent traffic assignment constraints.

3. THREE PERSPECTIVES ON NETWORK DESIGN

This study analyzes the CNDP-T from three perspectives: users, private toll road operators, and the government. Their considerations are discussed below:

1. Users are concerned with travel costs.

Users on each OD pair are concerned with their travel costs π_τ^{rs}. These travel costs are directly related to CS_τ^{rs}. The lower is the travel cost, the higher is the corresponding CS_τ^{rs}. To describe the change in travel cost "before" and "after" the improvement, we define the percentage decrease in OD travel cost of OD pair rs in period τ as follows:

$$PCC_{\tau}^{rs} = \frac{\pi_{\tau}^{rs,a} - \pi_{\tau}^{rs,b}}{\pi_{\tau}^{rs,b}} \times 100\% . \tag{22}$$

The superscripts b, a , respectively, refer to the cases of "before" and "after" the improvement.

2. Private toll road operators are concerned with their revenues.

The total toll revenue R_a on link a over the planning horizon is defined as summing the product of the toll and the volume on that link over time. Mathematically, the total toll revenue R_a on link a is expressed as:

$$R_a = \sum_{\tau} \rho_{a,\tau} v_{a,\tau} . \tag{23}$$

We define two related measures, the change in toll revenue CR_a and the percentage change in toll revenue PCR_a, to capture the change in toll revenue before and after the improvement. The change in revenue is written as:

$$CR_a = R_a^a - R_a^b . \tag{24}$$

The percentage change in revenue is expressed as:

$$PCR_a = \frac{R_a^a - R_a^b}{R_a^b} \times 100\% . \tag{25}$$

3. The government is concerned with the cost-effectiveness of the design.

We define the cost-effectiveness measure (CE) to be the ratio of the change in TCS to the total budget. It measures the return in TCS (a monetary term) as a result of the investment for network improvements. The design is cost-effective if the ratio is greater than 1. This measure also reflects whether TCS increases after the network improvement. When the ratio is greater than zero, TCS increases after the improvement. Mathematically, the cost-effectiveness measure can be stated as:

$$CE = \frac{TCS^a - TCS^b}{\sum_{\tau} B_{\tau}} . \tag{26}$$

4. NUMERICAL STUDIES

To study the budget sensitivity from the perspectives of users, private toll road operators, and the government in the CNDP-T, we set up three scenarios. The first scenario is to demonstrate the equity issue among the different user groups. The second scenario is to illustrate the equity issue between two private toll road operators. The last scenario is to study the cost-effectiveness issue.

Scenario 1: Equity on users

This scenario is to demonstrate the equity issue among the different user groups in the time-dependent network design problem. For ease of result exposition, we adopt a simple but actual network in Hong Kong, even though the formulation is applicable to more complex general networks. The network, as shown in Figure 1, is situated in a northwestern suburb of Hong Kong, consisting of 5 nodes, 6 links and 3 OD pairs. Links 1 and 2 form Tuen Mun Highway whereas Links 3 and 4 form Yuen Long Highway. Links 5 and 6 correspond to Route 3 (R3) and Route 10 (R10), which are toll roads. The three OD pairs are from Ting Kau to Tuen Mun, from Tuen Mun to Yuen Long, and from Ting Kau to Yuen Long. OD pairs (1,2) and (2,3) have one path while OD pair (2,3) has three paths. Five paths on this network are shown in Table 1. In this network, only links 1 and 4 can be widened. The parameters in this scenario include:

a) Link capacities: $c_1 = c_2 = c_5 = 3240$ vph, $c_3 = c_4 = 2160$ vph
b) Free flow travel times: $t_1^0 = 10$ min, $t_2^0 = t_3^0 = t_4^0 = 20$ min, $t_5^0 = t_6^0 = 15$ min
c) Maximum allowable link capacities: $u_1 = u_4 = 40000$ vph, $u_2 = u_3 = u_5 = u_6 = 0$ vph
d) Potential demands at period 1: $\tilde{q}_1^{12} = 9000$, $\tilde{q}_1^{13} = 15000$, $\tilde{q}_1^{23} = 12000$
e) Growth rates: $h^{12} = h^{23} = 0.10$, $h^{13} = 0.05$
f) Link performance function parameters: $\alpha = 0.15$, $\beta = 4$
g) Parameters of the travel demand functions: $\gamma^{12} = 1000$, $\gamma^{13} = 1500$, $\gamma^{23} = 800$
h) Value of time: ψ = HK$40/h
i) Converting factor: $n = 219000$
j) Convergence tolerance: $\varepsilon = 0.0001$
k) Construction cost parameter: $k = 1\times10^7$
l) Tolls: $\rho_{5,\tau}$=HK$9, $\rho_{6,\tau}$=HK$15
m) Length of each period: 5
n) Planning horizon: $[0,10]$

These values are chosen for the illustrative purposes only.

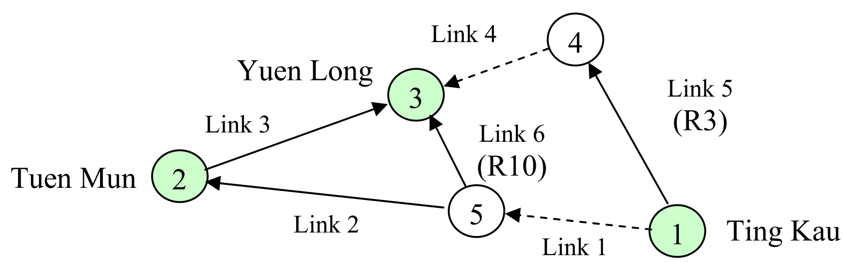

Figure 1. The Example Network

Table 1. Five Paths on the Network

OD pair	Path number	Node sequence	Minimum path time (min)
(1,2)	1	1-5-2	30
(2,3)	2	2-3	20
(1,3)	3	1-5-2-3	50
(1,3)	4	1-5-3	25
(1,3)	5	1-4-3	35

To show how well desired user equilibrium conditions are met under the stopping criterion ε = 0.0001, Tables 2 and 3 show the path flows and the corresponding path travel costs over the 10-year planning horizon when the initial budget equals HK$ 2×10^9. One can verify that in each period, all the used paths attain the minimum path travel cost whereas the unused one (i.e. path 3) has equal or higher path travel cost. That is, in each period, the deterministic user equilibrium conditions are satisfied.

Table 2. Path Flows

	Path number				
Period	1	2	3	4	5
1	2723.4	4005.8	0.0	4091.1	3473.2
2	2901.3	4836.3	0.0	5416.4	4415.1

Table 3. Path Travel Costs

	Path number				
Period	1	2	3	4	5
1	28.1	37.0	65.1	42.6	42.6
2	37.1	63.6	100.7	59.2	59.2

Figure 2 shows the capacity enhancements over time under a range of initial budgets. The zero-budget situation is equivalent to the before-improvement situation. All the improvements are done in the first period. In fact, this ought to be the best strategy as improvements completed within the first period can benefit both the first and second periods. Therefore, there is no reason to postpone the improvements to a future time. One would not expect this result, however, if the formulation also considers the available improvement budget in each period, the maintenance costs for the widened networks, and the discount and interest rates.

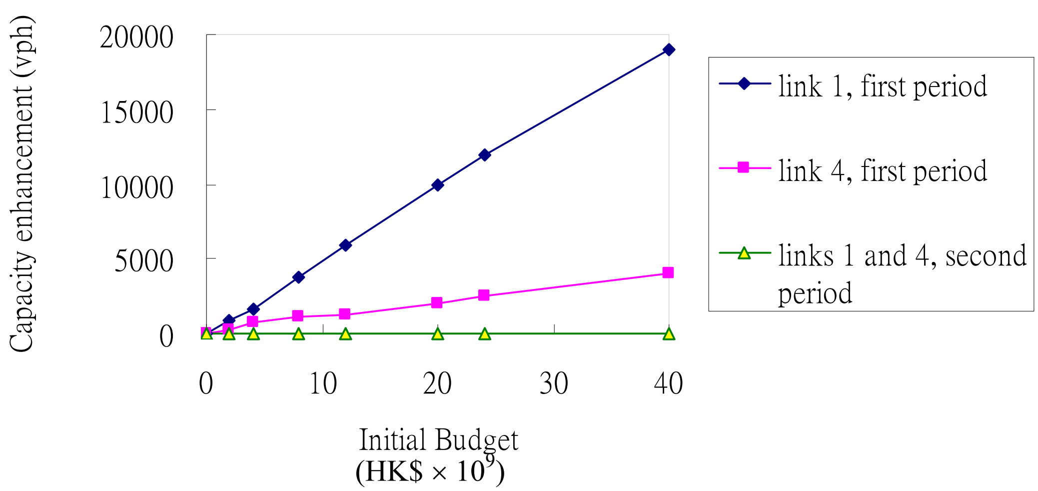

Figure 2. Capacity enhancements under a range of budgets

Another observation is that for a fixed budget, one would invest more on link 1 than on link 4 as improving link 1 can benefit the users between OD pairs (1,2) and (1,3) but improving on link 4 can only benefit the users between OD pair (1,3).

Figure 3 shows the time-dependent OD travel costs, which clearly illustrates the effects of the capacity enhancements and the increases in potential demands on the OD travel costs. In general, the travel costs in the second period are higher than that in the first period due to the large increases in potential demands. In addition, due to the capacity enhancements on links 1 and 4, the travel costs between OD pairs (1,2) and (1,3) decrease with increases in the budget.

In the extreme case, the OD travel costs remain constant. It is because when the budget tends to infinity, the capacities of links 1 and 4 tend to infinity and their travel times (costs) tend to their corresponding free-flow travel times (costs). However, the travel costs between OD pair (2,3) do not benefit from the improvements, because as travelers on this OD pair do not use links 1 and 4. They can only be benefited when link 3 is widened or the congestion on link 3 is alleviated. The percentage decreases of OD travel costs are also not the same (Figure 4), meaning that the travelers between these OD pairs receiving benefit to different extents. This raises the issue of equity - *network improvements that maximize the overall consumer surplus do not necessarily benefit all OD pairs to the same extent.*

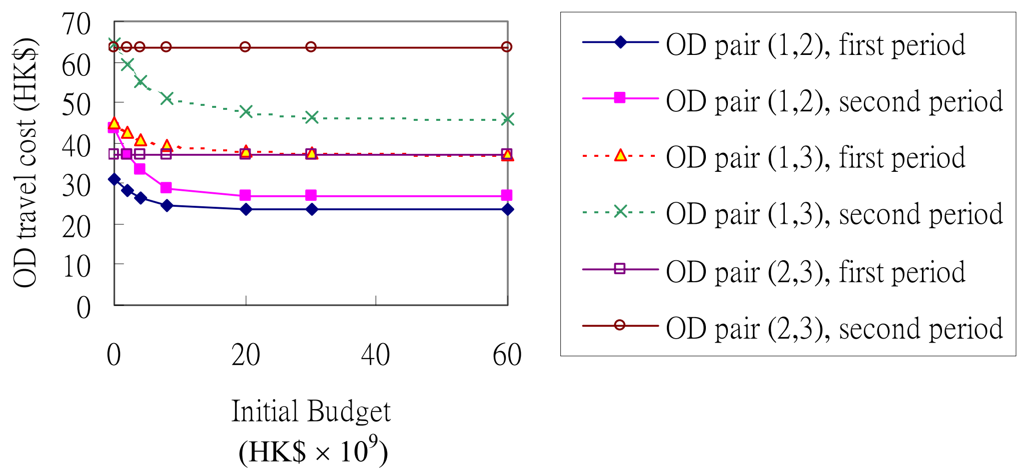

Figure 3. Time-dependent OD travel costs under a range of budgets

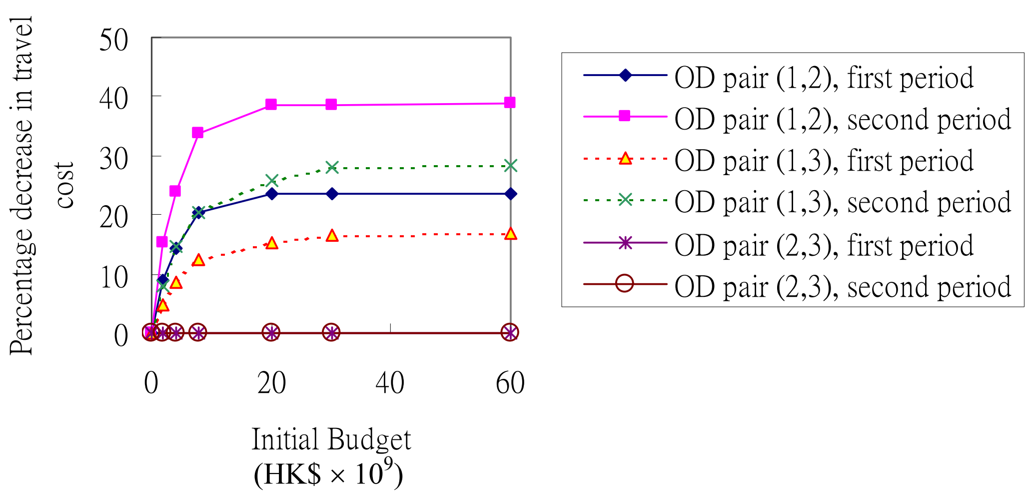

Figure 4. The percentage decrease in OD travel cost under a range of budgets

In this scenario, capacity enhancements do not increase the OD travel costs. Nonetheless, travel costs (and hence the corresponding consumer surpluses) of some OD pairs could decrease despite network improvements (see Lo and Szeto, 2003). This is an extreme case of the equity issue - some OD pairs may be impaired by the network improvement whereas others benefit from it.

Scenario 2: Equity on toll road operators
This scenario is to illustrate the equity issue on different toll road operators due to capacity enhancements. We adopt all the settings in Scenario 1 and plot the change in toll revenue of each private operator over time against the budget as shown in Figure 5. R3 and R10 correspond to Route 3 and Route 10. According to this figure, as the budget increases, the change in toll revenue of each operator increases monotonically up to its respective constant value, when the improvement is so big that all the demand that can be induced is induced. This of course, represents only an extreme scenario.
In terms of the change in toll revenue (Figure 5), the toll revenue for R3 is larger than that for

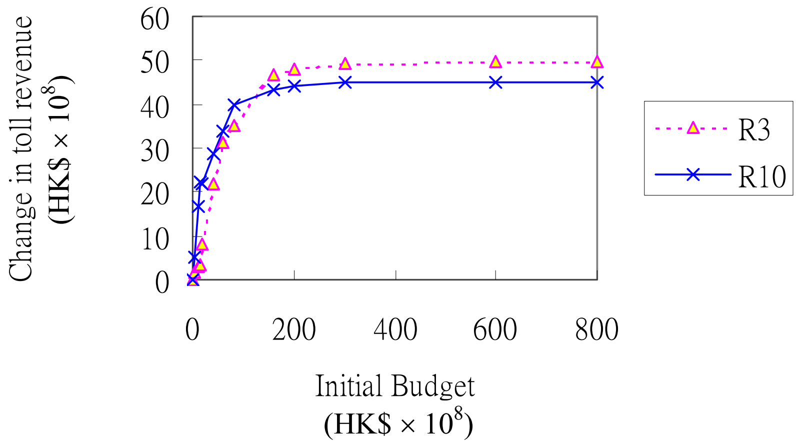

Figure 5. The change in revenue under a range of budgets

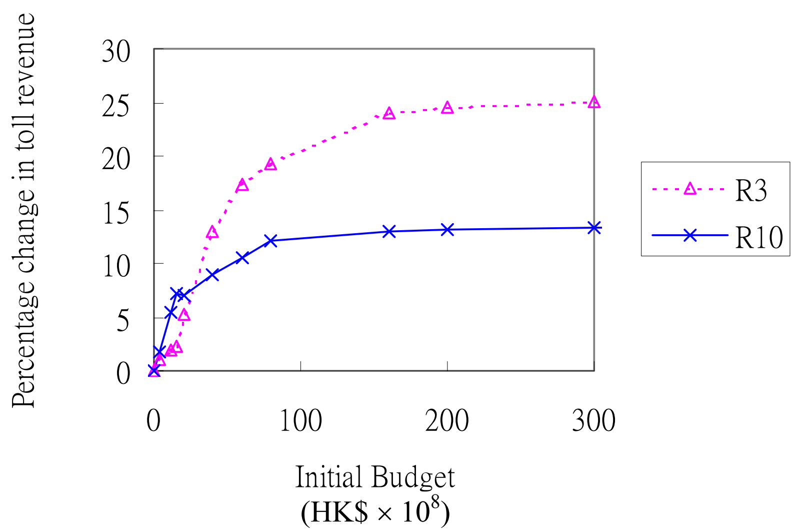

Figure 6. The percentage change in revenue under a range of budgets

R10 when the budget is larger than HK\$ 1.2×10^{10}, despite that the toll on R10 is higher. Outside this budget range, the reverse is true. This shows that *capacity improvements can lead to unequal benefit distributions on each toll road operator*. In terms of the percentage change in revenue, Figure 6 illustrates a similar pattern. In this scenario, capacity enhancements increase the revenues of both operators. However, conceivably, one can think of scenarios wherein network improvements could hurt the toll revenues. This would happen for improvements on a toll-free facility that runs in parallel to the toll facility. The actual outcome depends on whether the toll and toll-free facilities are complementary to each other (as in this numerical example) or substitutes of each other (as in the latter case discussed but not illustrated).

Scenario 3: Cost- effectiveness of the design from the government perspective
This scenario studies the cost-effectiveness of a design under a limited budget. Scenario 3 adopts the same set of parameters as in Scenario 1, except that we consider two cases, with different values on the construction cost parameter k for each case. Case 1 is the equivalent to scenario 1 ($k = 1\times10^7$). Case 2 is obtained by increasing the construction cost parameter by a factor of 10 ($k = 1\times10^8$). The results are shown in Figure 7. The x-axis refers to the budget whereas the y-axis refers to the change in TCS. The dashed line represents the 45-degree line. The upper curve is for case 1 ($k = 1\times10^7$) and the lower curve is for case 2 ($k = 1\times10^8$).

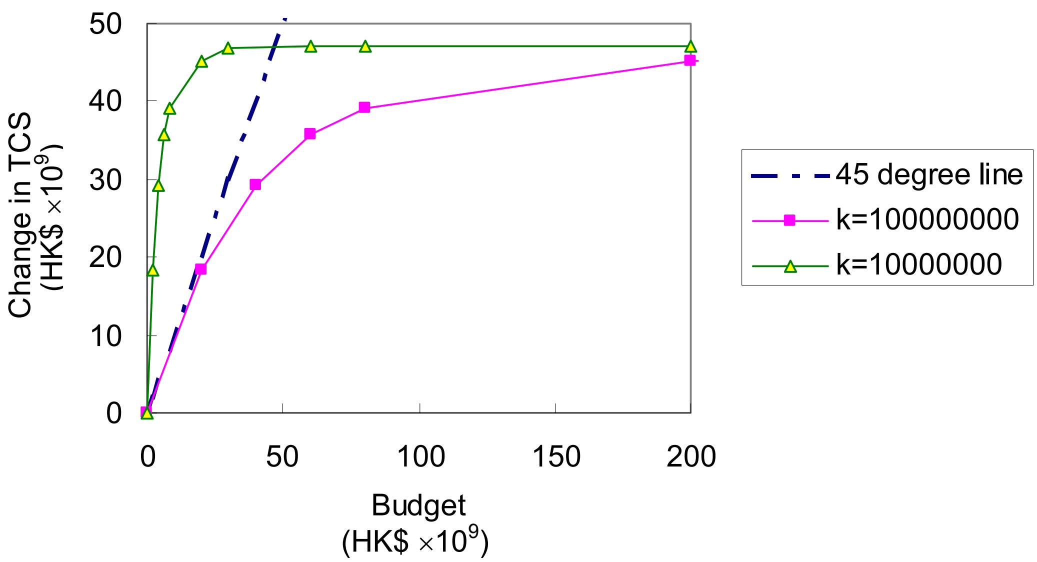

Figure 7. The change in the total consumer surplus against the budget.

Case 1: $k = 1\times10^7$

The upper curve shows the case of diminishing return. It goes up first, levels off gradually, and eventually remains constant at the level of HK$\$4.7\times10^{10}$. As explained before, as the budget goes to a very high level, the improved links attain their free flow travel costs, with the corresponding CS approaches a constant value.

The upper curve is above the 45-degree line when the budget is between HK$0 and HK$\$4.6\times10^{10}$. This implies that the ratio of the change in TCS to the budget is large than 1 within this budget range, or that each dollar of investment on the network results in more than one dollar of return in CS. Nevertheless, when the budget is larger than HK$\$4.6\times10^{10}$, the ratio is smaller than 1, implying that it is not cost-effective to invest beyond this range.

Case 2: $k = 1\times10^8$ (increased by a factor of ten)

Similar to case 1, the curve shows the case of diminishing return. However, the lower curve is always lower than the 45-degree line but above the x-axis. This means that the ratio is always less than 1 but greater than zero for all budget values. From a cost-effectiveness point of view, the network improvement is not worthwhile. Nevertheless, each dollar spent in the network improvement does introduce an increase in CS.

By comparing the above three scenarios, we find all-win situations cannot happen when the budget is larger than HK$ 4.6×10^{10} and the construction cost parameter k equals 1×10^{7}. Careful considerations must be included in planning network improvements over time so as to strike a balance in benefit distributions among the involved parties.

4. CONCLUDING REMARKS

Designing transport network improvements often involve three parties: network users, private toll road operators, and the government. Each of these parties has distinctive objectives. By extending the CNDP-T formulation proposed by Lo and Szeto (2003), this study develops a single-level optimization program to plan the network improvements over time and to study the tradeoff among these parties under a range of budgets. Three numerical examples are provided to demonstrate the equity issues among users and among private toll road operators, and the cost-effectiveness of the design from the government perspective. The results also show that there is a tradeoff among the three parties.

The results presented here are somewhat preliminary. Nevertheless, we believe the discussions have brought up a number of research extensions. Firstly, network improvements are often subject to time-dependent tolls, annual budgets, economies of scale, the associated maintenance costs, and the tradeoff between inflation and interest rates, and the discount rate of benefits. All of these factors should be duly considered for a complete analysis. Secondly, the CNDP-T is non-convex due to the user-equilibrium constraints. It is beneficial to develop effective sensitivity-based or penalty-based algorithms to deal with these constraints for larger-scale implementations. Thirdly, deterministic user equilibrium may not be a realistic assignment principle. Using the formulation developed herein as a platform, one may extend it for other types of assignment principles, such as the stochastic dynamic user optimal principle (Ran and Boyce, 1996), the probabilistic user equilibrium principle (Lo and Tung, 2003), or others. Finally, the proposed model is only applicable to the situation when the stated assumptions in the text are nearly satisfied including that the demand structure and financial environment do not substantially differ from the ones predicted. The uncertainties in the demand structure and financial environment are not considered in this paper. They are incidental and the most important aspect in the risk management analysis. Indeed, one important research direction is to capture these uncertainties in the planning process.

ACKNOWLEDGMENT

The authors thank the constructive comments of the referees. This research is sponsored by the Competitive Earmarked Research Grant HKUST6154/03E from the Hong Kong Research Grant Council.

REFERENCES

Abadie, J. and J. Carpentier (1969) Generalization of the Wolfe reduced gradient method to the case of nonlinear constraints, in R. Fletcher (ed.), **Optimization**, New York: Academic Press, pp. 37-47.

Friesz, T.L., G. Anandalingam, N.J. Mehta, K. Nam, S.J. Shah and R.L. Tobin (1993) The multiobjective equilibrium network design problem revisited: a simulated annealing approach, **European Journal of Operational Research, Vol. 65, No. 1**, 44-57.

Gartner, N.H. and C. Stamatiadis (2001) Combining traffic assignment and adaptive control in a dynamic traffic management system, in E. Schnieder and U. Becker (eds), **Control in Transportation Systems 2000: Proceedings volume from the 9th IFAC Symposium, vol.1**. UK: Elsevier Science, pp. 281-286.

Lam, W.H.K., M.L. Tam, and M.G.H. Bell (2002) Optimal road tolls and parking charges for balancing the demand and supply of road transport facilities, in M.A.P Taylor, **Transportation and Traffic Theory in the 21st Century: proceedings of the 15th International Symposium on Transportation and Traffic Theory**, New York: Elsevier Science, pp. 561-582.

Lo, H. and W.Y. Szeto (2002) A cell-based variational inequality formulation of the dynamic user optimal assignment problem, **Transportation Research, Vol. 36B, No. 5**, 421-443.

Lo, H. and W.Y. Szeto (2003) Planning transport network improvements over time, In D.H. Lee (ed), **Urban and Regional Transportation Modeling: Essays in Honor of David Boyce**. In press.

Lo, H. and Y.K. Tung (2003) Network with degradable links: capacity analysis and design, **Transportation Research, Vol. 37B, No. 4**, 345-363.

Maher, M.J., X. Zhang and D. Van Vliet (2001) A bi-level programming approach for trip matrix estimation and traffic control problems with stochastic user equilibrium link flows, **Transportation Research, Vol. 35B, No. 1**, 23-40.

Meng, Q., H. Yang and M.G.H. Bell (2001) An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem, **Transportation Research, Vol. 35B, No. 1**, 83-105.

Ran, B. and D. Boyce (1996) **Modeling Dynamic Transportation Networks: an Intelligent Transportation System Oriented Approach**. Berlin: Springer.

Wardrop, J. (1952) Some theoretical aspects of road traffic research, **Proceedings of the Institute of Civil Engineers, Part II**, 325-378.

Wong, S.C. and H. Yang (1997) Reserve capacity of a signal-controlled road network, **Transportation Research, Vol. 31B, No. 5**, 397-402.

Yang, H. and M.G.H. Bell (1998) Models and algorithms for road network design: a review and some new developments, **Transport Reviews, Vol. 18, No. 3**, 257-278.