Article:

Shepherd, S.P. and Sumalee, S. (2004) A genetic algorithm based approach to optimal toll level and location problems. Networks and Spatial Economics, 4 (2). pp. 161-179. ISSN 1572-9427

https://doi.org/10.1023/b:nets.0000027771.13826.3a

Published paper

Shepherd, S.P. and Sumalee, A. (2004) A Genetic Algorithm Based Approach to Optimal Toll Level and Location Problems. Networks and Spatial Economics 4(2), pp.161-179

A Genetic Algorithm Based Approach to Optimal Toll Level and Location Problems

SIMON SHEPHERD

AGACHAI SUMALEE[*]

Institute for Transport Studies, University of Leeds, LS2 9JT, UK

Abstract

The derivative based approach to solve the optimal toll problem is demonstrated in this paper for a medium scale network. It is shown that although the method works for most small problems with only a few links tolled, it fails to "converge" for larger scale problems. This failure led to the development of an alternative genetic algorithm (GA) based approach for finding optimal toll levels for a given set of chargeable links. A variation on the GA based approach is used to identify the best toll locations making use of "location indices" suggested by Verhoef (2002).

Keywords: Bilevel optimisation, second-best tolls, optimal tolls, optimal location, genetic algorithms

Introduction

The concept of road pricing emerged from the idea that the cost paid by the road user (called marginal private cost or perceived cost) is actually lower than the actual cost (s)he imposes (called marginal social cost) (Pigou 1920; Knight, 1924; Walters, 1961; and Vickrey, 1969). The development of the theory of marginal cost pricing relies heavily on the assumption of first-best conditions. The assumptions of the first best condition are not usually satisfied (see, Sharp 1966). These assumptions are for example that tolls are imposed on all links or that prices in other modes are all controlled in a first best manner. The main area of research has been the second-best condition of marginal cost

[*] Author for correspondent, email: asumalee@its.leeds.ac.uk

tolling where not all links in a network can be tolled (Levy-Lambert, 1968; Marchand, 1968; Arnott et al, 1990; Glazer and Nikanen, 1992; Liu and McDonald, 1999; Small and Yan, 2001).

In the optimisation context, the second-best marginal cost toll problem is categorised as a mathematical programming problem with equilibrium constraints (MPEC), a special case of the bi-level optimisation programming problem (BLPP). The regulator tries to set the toll locations and toll levels to optimise his or her objective whilst the users attempt to minimise their own travel costs. The optimal toll problem can also be seen as a special case of the network design problem. Several methods have been proposed to tackle this challenging problem. Most procedures to solve the BLPP use derivatives and can be put into one of the following categories, heuristic iterative optimisation method (Steenbrink, 1974; Allsop, 1974; Suwansirikul et al; 1987), transforming the BLPP to a single level optimisation program, linearisation method (LeBlanc and Boyce, 1986; and Ben-Ayed et al, 1988), and stochastic search methods[1] (Friesz et al, 1992; Cree et al, 1998; Yang et al, 2000; and May et al, 2002). There exists a diverse range of techniques used to transform the BLPP to a single level optimisation program. These include sensitivity based analysis (Tobin and Friesz, 1983; Friesz et al, 1990;Yang, 1997), Karush-Kuhn-Tucker (KKT) based method (Marcotte, 1983; Marcotte, 1986; and Verhoef, 2002), using the system optimal solution to formulate the set of tolls for the second-best case under user equilibrium (Bergendorff et al, 1996; Hearn and Ramana, 1998; and Hearn and Yildirim, 2002), and a marginal function based method (Meng et al, 2001).

The application of GA in the optimal toll design problem is not new. As mentioned earlier, Cree et al (1998) developed the GA based method to solve the optimal toll problem but not the location problem. Yang et al (2002) went one step further by using GA to find the optimal closed cordon but with a uniform toll level only. The originality of our GA based method is indeed the linkage between GA and location indices. The toll levels of the optimal combination of tolled links are allowed to be varied which is also considered to be a new development. Our paper also raises the question of the practicality of the derivative-based approach for solving the optimal toll problem with a large-scale

[1] e.g. simulated annealing and genetic algorithms

network. This problem is not related to the existing problematic characteristics of MPEC or BLPP (e.g. non-convex feasible set or non-smooth objective function); rather it is related to the fact that the theoretical development of derivative-based methods always assumes perfect convergence of the user equilibrium condition whereas this may not be the case for a large scale network solved using an iterative assignment procedure. We show that when the convergence of user equilibrium is not perfect the derivative-based approach may even converge to a sub-optimal point.

This paper firstly demonstrates a derivative based method proposed by Verhoef (2002) to solve the optimal toll level and location problems and discusses problems that can occur. New methods are then developed to tackle the problems based on the use of genetic algorithms (GA). The paper consists of four further sections. The next section shows the formulation of the optimal toll problem as a MPEC and explains briefly the derivative based method. Then, it discusses circumstances which cause the derivative based method to fail. Similarly, the method to find the optimal toll location is investigated and a weakness is identified. Section three then proposes three new methods based on GA. Section four displays numerical results comparing the GA and derivative based approaches. The final section draws conclusions.

1. Derivative based method to optimal toll design and its drawbacks

1.1 Optimal toll level on a specified set of links[2]

The problem of defining the optimal toll level on a specified set of tolled links (termed OPT1) can be formulated as a MPEC. The objective function of this optimisation program is to maximise the social welfare function following Marshallian's rule measure which can be formulated as:

$$W\left(\mathbf{F},\tau\right)=\sum_i\int_0^{T_i}D_i(x)dx-\sum_j\sum_p\delta_{jp}\cdot F_p\cdot c_j \tag{1.1}$$

[2] Notation is given in appendix A.

where **F** is a vector of path flows. Once the tolls are implemented, road users will respond to the tolls by changing their routes or deciding not to travel. This response is captured by assuming the road users behave according to Wardrop's user equilibrium rule (Wardrop, 1952). This condition is the so called user equilibrium condition (UE). The UE can be formulated as a complementarity slackness condition expressing the stationary point of the solution toward the optimisation problem (Smith, 1979). The complementarity constraint (CP) condition of UE with elastic assignment is as follows:

$$0 \le F_p \perp \left(\sum_j \delta_{jp} \cdot \left(c_j + \varepsilon_j \cdot \tau_j \right) - D_i \right) \ge 0 \qquad \forall p \in \Pi \tag{1.2}$$

The CP constraint as appeared in (1.2) imposes several problematic conditions to the optimisation program, i.e. disjunctive characteristic of constraints and non-convex constraint. In order to bypass this difficulty, Verhoef (2002) assumed that only used paths $p \in \Pi$ are included in the optimisation problem which makes the complementarity slackness condition above reduce to:

$$\left(\sum_j \delta_{jp} \cdot \left(c_j + \varepsilon_j \cdot \tau_j \right) - D_i \right) = 0 \qquad \forall p \in \left\{ p \middle| p \in \Pi \text{ and } F_p > 0 \right\} \tag{1.3}$$

We then obtain a single level optimisation program of optimal toll level as follows:

$$\begin{aligned} &\max W\left(\mathbf{F}, \tau\right) = \sum_i \int_0^{T_i} D_i(x)dx - \sum_j \sum_p \delta_{jp} \cdot F_p \cdot c_j \\ &s.t. \\ &\left(\sum_j \delta_{jp} \cdot \left(c_j + \varepsilon_j \cdot \tau_j \right) - D_i \right) = 0 \qquad \forall p \in \left\{ p \middle| p \in \Pi \text{ and } F_p > 0 \right\} \\ &g_l\left(\mathbf{F}\right) \ge 0 \qquad l \in \left\{1, ..., L\right\} \end{aligned} \tag{1.4}$$

where g_i(**F**) is the function ensuring the flow conservation and non-negativity requirement. Verhoef (2002) proposed the method to solve the OPT1 by maximising the following Lagrangian:

$$\Lambda = \sum_i \int_0^{T_i} D_i(x)dx - \sum_j \sum_p \delta_{jp} \cdot F_p \cdot c_j + \sum_p \lambda_p \cdot \left(\sum_j \left(\delta_{jp} \cdot \left(c_j + \varepsilon_j \cdot f_j \right) - D_i \right) \right) \tag{1.5}$$

subject to the set of feasible path flows and non-negative path flows. The first two terms represent the Marshallian measure of social welfare as shown in (1.1). The λ_p are the Lagrange multipliers associated with each used path under the equilibrium condition. The third term associated with the Lagrange multiplier is the complementarity slackness constraint for the user equilibrium condition. Note that only the used paths under equilibrium condition ($F_p > 0$) are included in the Lagrangian in order to reduce the CP constraints to the normal equality constraints. Verhoef (2002) derived the first order condition of this Lagrangian. Shepherd et al (2001a) utilised this first order condition to develop the mathematical program, termed CORDON, linked with SATURN (Van Vliet et al, 1982) to solve OPT1.

1.2 Performance of CORDON and the effect of assignment convergence error

The main concern with the method to solve the optimal toll level, termed CORDON, is the assumption regarding the set of used paths. The relaxation of the complementarity slackness condition to include only the set of used paths ($F_p > 0$) allows the introduction of Lagrange multipliers. However, the use of these multipliers relies on a perfectly converged set of paths i.e. each path must satisfy the equilibrium condition such that the minimum O-D costs and path costs are equal. It is not a trivial task to seek a perfectly converged solution for user equilibrium conditions in medium and large-scale networks. Indeed SATURN currently uses the Frank-Wolfe algorithm to solve the traffic assignment problem, which is an iterative process and as such is stopped when certain convergence criteria are satisfied. In terms of traffic assignment the solutions presented in this paper would be considered well converged, however the effects on the lagrangian can, as we demonstrate, be significant as the assignment errors are magnified by the lagrange multipliers.

Figure 1 shows the network used to discuss the CORDON method and possible problems. There are 18 links which can be considered in turn as a single toll point. Of these, five links require a negative toll to increase the total welfare, these are the four bypass links (6-4, 4-6, 3-6, 6-3) and link (4-2).

This is a reasonable result, as the flow-delay parameters are such that re-routing to these links would provide benefits for the system as a whole. However the SATURN program requires a positive toll as input and although the optimisation process points towards a negative toll we have limited the predicted toll to be positive – thus the process is considered to have selected the correct solution here.

Of the remaining 13 links 8 can be solved by the proposed method. An example of a well-behaved solution is shown for link 4-5 in figure 2. The "dots" represent the iterations required to locate the optimum charge level which maximises the total benefit or social welfare function defined earlier. The remaining 5 links exhibit multiple local optima as shown in figure 3 for link 1-2. Here it can be seen that the optimisation process finds the first local optimum. The double hump is a result of the interactions between route choice and between OD pairs which was limited in Verhoef's ten link example. The global optimum can be found if the starting point is changed from no toll to a toll of around 500 seconds, but it is not obvious how this might be implemented in general other than by some general perturbation approach. Verhoef (2002) previously claimed that multiple optima would be a rare commodity, but as can be seen here there are 5 examples from only 18 links.

Figure 4 shows the benefit curve and the CORDON process for link 2-4. The process converged to a point which is neither a local optimum nor a deflection point. The CORDON process uses the first order derivatives of the lagrangian shown in equation (1.1). In the case where all the paths included are in perfect equilibrium, the Lagrangian matches the objective function (social welfare function), as the last term which includes the impact of the convergence error reduces to zero. However, if an assignment error occurs in SATURN, this can cause a discrepancy between the true objective function and the Lagrangian. The assignment convergence errors are magnified by the Lagrange multipliers and then added to the real objective function in the Lagrangian causing the difference between the real benefit (social welfare improvement) and the Lagrangian value.

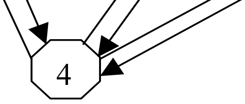


Figure 1: A theoretical network

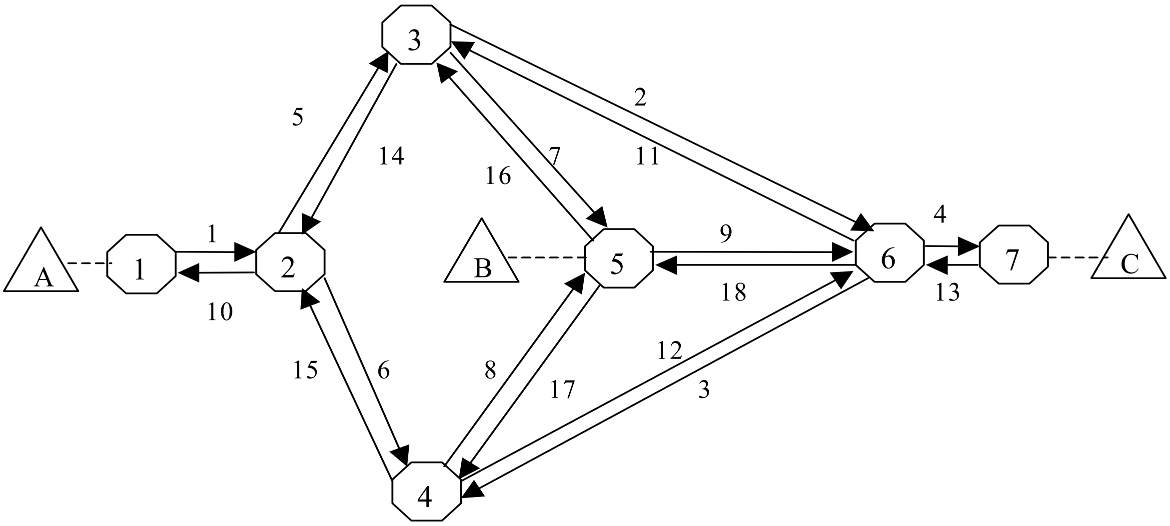


Figure 2 : Benefit versus charge on link 4-5.

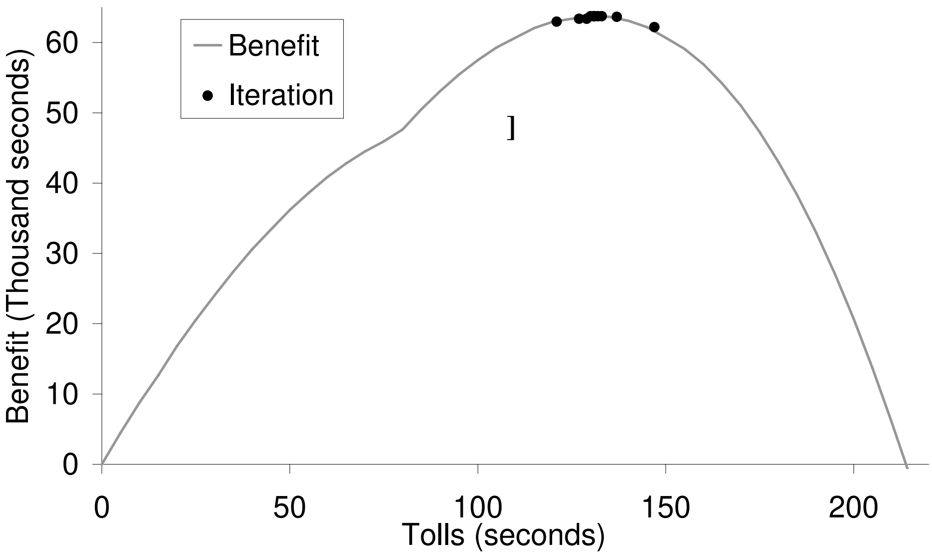


Figure 3 : Benefit versus charge on link 1-2

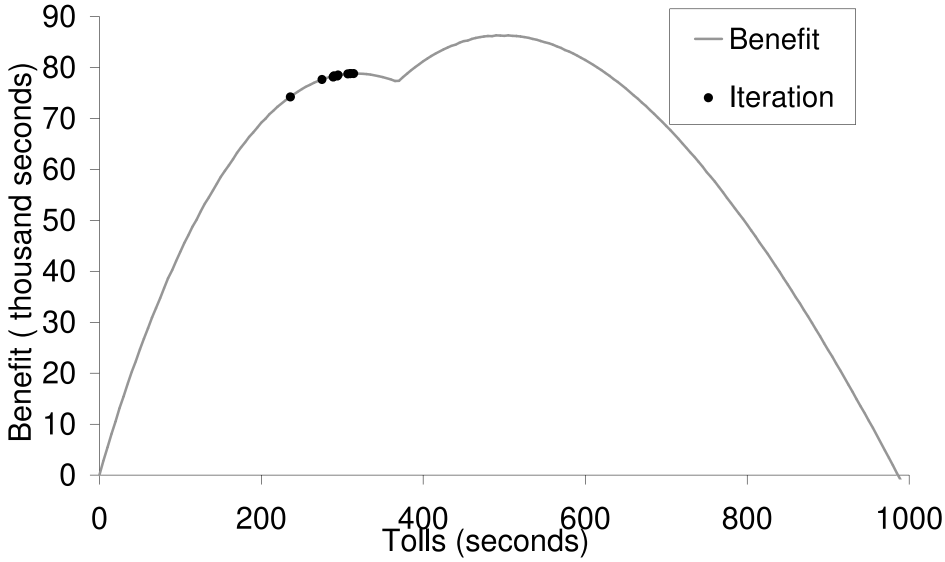


Figure 5 shows the comparison between the real total benefit and the Lagrangian as the charge is varied on link 2-4. The CORDON process converges to the local optimum of the Lagrangian resulting in a charge of 27 seconds. Note that for charges in the range 10-40 seconds there is always a convergence error. Closer investigation showed that this occurred as the demand for a particular path was reduced and eventually set to zero thus changing the path set. As the charge level is increased there are fewer occasions where the convergence error is a problem and the probability that an iteration of the CORDON process hits the convergence problem is reduced significantly.

Figure 4: Benefit versus charge on link 2-4

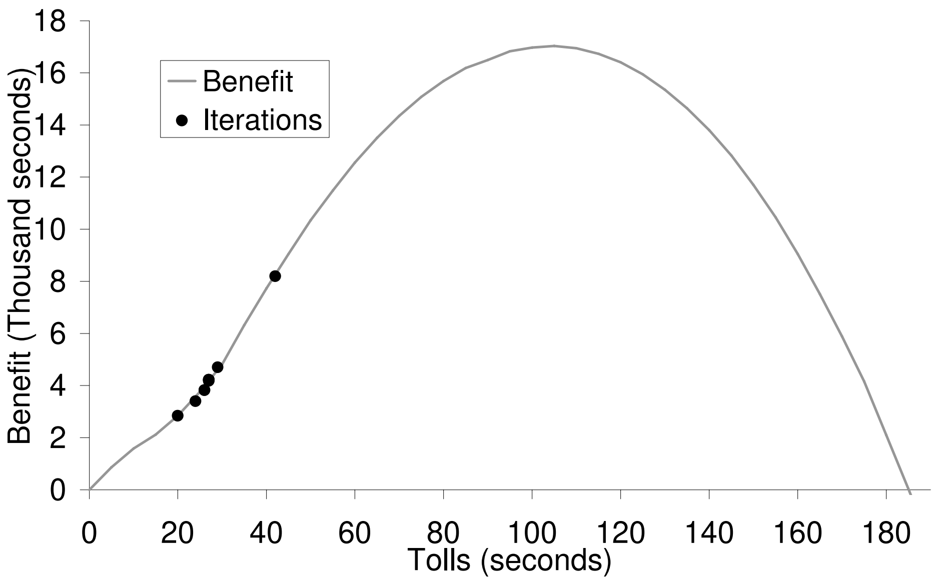
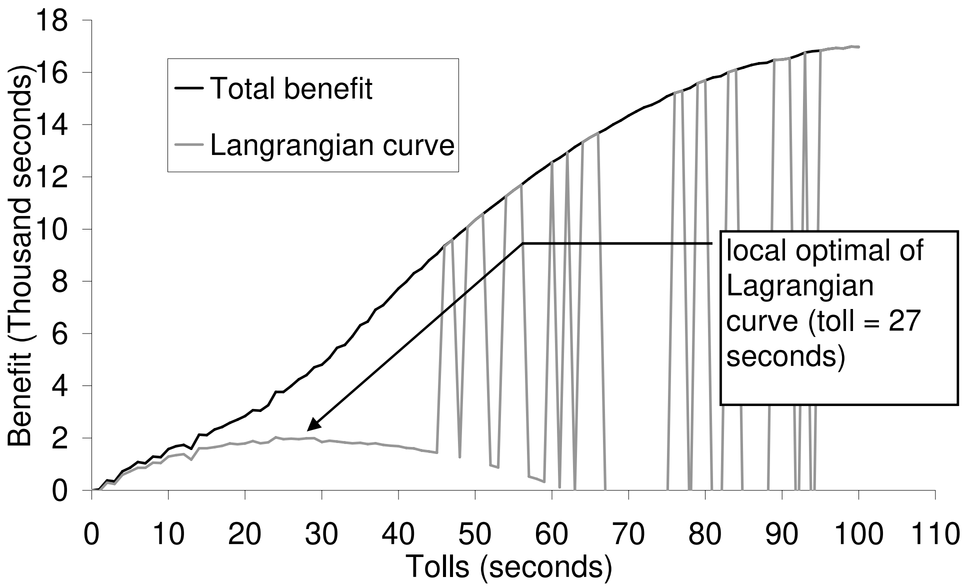


Figure 5: Lagrangian and total benefit curve link 2-4



Thus, we have demonstrated the drawbacks of the derivative based method. The possibility of multiple optima and lack of perfect convergence from the assignment algorithm mean that the CORDON process is unable to find the global optimum in all cases. The requirement for a perfectly converged traffic assignment implies the need for a better traffic assignment algorithm for a large scale problem. In Section 2, the genetic algorithm based approach to solve the OPT1 problem is presented as an alternative to the derivative based method.

1.3 Optimal toll location in a general network

The other problem when designing a toll based road pricing system is to define the optimal toll point locations given the desired number of tolled points (termed OPT2). A related problem is to find the optimal number of tolled points and locations simultaneously while considering implementation costs (termed OPT3). The most straightforward approach to solve the optimal toll location problem is to test all the combinations of tolled points. However, this would be computationally demanding due to the massive number of possible combinations of tolled points. Suppose that we try to choose the best t tolled links from the j links in the network, the number of all possible combinations of the tolled links will be $\frac{j!}{t!(j-t)!}$. Verhoef (2002) proposed an incremental approach which Shepherd et al (2001b) adapted and termed LOCATE.

The LOCATE process is an extension of CORDON and involves building up a list of toll points incrementally, by choosing links one by one on the basis of a location index. The location indices are the approximation of the welfare gains that would result from placing optimal charges in particular locations. They use the predicted toll from the first iteration of the CORDON process combined with the shadow prices associated with the link(s) considered. Although previously selected toll points are always included, the charge levels are allowed to vary each time an additional link is added. Those interested in the details of the CORDON process and the LOCATE process should refer to Shepherd et al (2001a) and Shepherd et al (2000b) respectively.

Shepherd et al (2001b) also showed that LOCATE can fail to identify the best pair of tolled links from a simple five-link network. This happened because the best “single” tolled link was not part of the best “pair”. To overcome this weakness Verhoef (2002) suggested a greedy search could be used. However, it is not practical to implement this strategy with large-scale networks due to the number of possible combinations mentioned earlier. Thus, the idea of genetic algorithms (GA) is adopted to generate combinations of tolled points instead. The next sections introduce the concept of GA based methods to solve the problems OPT1, OPT2, and OPT3.

2. Genetic algorithm based methods

2.1 Introduction to genetic algorithms

Genetic algorithms (GA) are one of the artificial intelligence exhaustive searching techniques; they are stochastic algorithms whose search methods model some natural phenomena: genetic inheritance and Darwinian strife for survival. Davis and Steenstrup (1987) stated that:

“The metaphor underlying genetic algorithms is that of natural evolution. In evolution, the problem each species faces is one of searching for beneficial adaptations to a complicated and changing environment .The ‘knowledge’ that each species has gained is embodied in the makeup of the chromosome of its members.”

The basic idea of the GA approach is to code the decision variables of the problem as a finite string (called ‘chromosome’) and calculate the fitness (objective function) of each string. Chromosomes with a high fitness level have a higher probability of survival. The surviving chromosomes then reproduce and form the chromosomes for the next generation through the ‘crossover’ and ‘mutation’ process. The method of GA is widely applied in many disciplines, but most applications have to modify the GA to the problem or change the problem to be compatible with GA. The main parts in the modification process are the design of chromosome encoding and of the genetic operators (crossover and mutation processes) in order to maintain the search within the feasible space. In the following sections, three methods are developed to solve the OPT1, OPT2 and OPT3 problems.

2.2 A method to solve optimal toll levels (GA-CHARGE)

The GA-CHARGE approach is developed to solve the OPT1 problem. The process of GA-CHARGE randomly generates an initial set of chromosomes representing possible combinations of charge levels on a predefined set of links. The benefits in terms of social welfare improvement are evaluated for each charge level by running SATURN. GA-CHARGE then selects the parent chromosomes for the next generation based on the performance of each chromosome. Since the fitness value in GA-CHARGE can be negative, the selection is based on the tournament selection process (Michalewicz, 1992). The genetic operators, crossover and mutation, are then randomly applied to the parents to produce the offspring.

2.2.1 Chromosome encoding

Let *t* be the number of predefined tolled links and let *r* be the predefined maximum toll level. Each chromosome represents a set of toll levels for the t-tolled links in binary format. The structure of the chromosome is therefore a matrix **A** with t columns and k rows where k is determined by the number of digits required to represent the maximum toll in binary format. Figure 6 shows an example chromosome (**A** matrix) for ten tolled links. The toll on each link is defined by the binary number in each column which is shown in the bottom row.

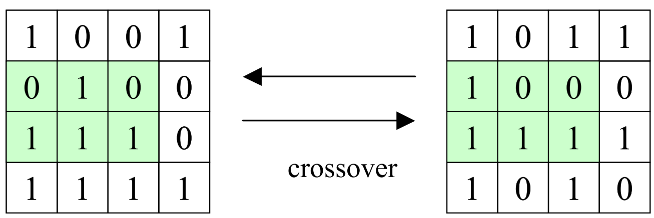
Figure 6: Chromosome structure for GA-CHARGE

$$\begin{array}{c} \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \\ \quad\;\; \begin{matrix} 31 & 21 & 3 & 16 & 11 & 20 & 15 & 10 & 19 & 22 \end{matrix} \end{array}$$

2.2.2 Crossover and mutation process

The crossover process is to select at random a partition from the chromosome matrix which is then switched between two "mated" chromosomes (See Figure 7).

Figure 7: An example of the crossover process in GA-CHARGE



After the crossover process, the mutation process is applied to the offspring. The mutation process randomly chooses cells to be "mutated". If selected, the value in that cell is changed from 0 to 1 or vice-versa.

2.3 A method to solve optimal toll location based on location indices (GA-LOCATE)

This section explains the approach to use GA to solve the optimal toll location problem, termed GA-LOCATE. The GA process is used to randomly generate and evolve the combinations of the tolled points (chromosome). The location index (see Section 1.3) of each combination is calculated and used as its fitness value. The selection process is based on "stochastic universal sampling" which uses a single wheel spin (Michalewicz, 1992). The so called "roulette wheel" is constructed where each slot represents a chromosome. The slots are sized according to the fitness of each chromosome. The size then represents the probability of a chromosome being selected.

2.3.1 Chromosome encoding

The user inputs the number of tolled points required. The adapted chromosome for OPT2 varies the length of chromosome to represent the required number of tolled links and each bit represents a selected link. A list of suitable chargeable links (called candidate links) can be prepared in advance to reduce the problem. With this structure, the length of the chromosome already controls the required number of tolled points. However, one problem with this structure is duplicating the selected links during the genetic operations.

2.3.2 Crossover and mutation process

Two points on the chromosome are randomly chosen and all "bits" in between these two points of the parents are switched. The mutation process is normally used to avoid the premature convergence of the GA process. In the traditional GA process, the value assigned to each bit is either 1 or 0 (as used in GA-CHARGE). In our modified chromosome, the possible value in each bit is the integer number between 1 and the highest number of candidate links.

2.4 A method to solve toll location problems with implementation costs (GA-LOCATEII)

The GA-LOCATEII process is developed to solve the problem OPT3. The general algorithm is similar to GA-LOCATE but includes implementation costs. The traditional chromosome structure of GA, string of binary bits, is adopted since there is no constraint on the number of toll points required in this problem. One column is required for each candidate link. If the value is 1, the corresponding link is to be tolled. The location index is calculated and the implementation and operation costs per toll point are subtracted. The standard crossover and mutation processes are also adopted. Note that since the fitness value (location indices net of costs) can be negative which causes a problem for the roulette wheel approach. Thus, the linear ranking approach proposed by Whitley (1989) linked with stochastic universal sampling is adopted. The slots in the roulette wheel are sized according to the chromosome at rank *i*, where the first is the best chromosome, by the following equation:

$$p_i = \frac{1}{\|P\|} \cdot \left(2 - c + (2c - 2) \cdot \left(\frac{\|P\| - i}{\|P\| - 1} \right) \right) \tag{2.1}$$

where $\|P\|$ is the size of the population set *P*, and $1 \leq c \leq 2$ is *"the selection bias"*: higher values of *c* cause the system to focus more on selecting only the better individuals. The best individual in the population is thus selected with the probability $\frac{c}{\|P\|}$; the worst individual is selected with the probability $\frac{2-c}{\|P\|}$.

3. Numerical results

3.1 Network description and experimental setting[3]

In this section, the methods developed in the previous sections are tested with a medium-scale network. Figure 8 shows the network which is based loosely on the City of Leeds network in the UK.

[3] A full detail of network description can be found at http://www.personal.leeds.ac.uk/~traas/NASE.htm

It should be noted, as the detail of the network has been reduced to decrease the complexity and computation time, the network cannot be considered as a comparative model to the real Leeds network. There are 89 directed links and 14 zones in this network. The triangular nodes represent the zones. The network is a "bufferised" version of a SATURN network, which means the supply is represented by independent flow-delay relationships for each link. The network is used for the following tests:-

i. Given three pre-defined charging cordons, the CORDON and GA-CHARGE processes are used to find the optimal toll levels around each cordon, see Figure 8 (OPT1);

ii. Given the desired number of tolled links, the LOCATE and GA-LOCATE processes are applied to find the optimal location of the tolled links (OPT2) and charge levels are then optimised using CORDON;

iii. Finally, GA-LOCATEII is applied to find the optimal number of the tolled points and the optimal tolled links assuming implementation costs (OPT3) .

Figure 8: MINILEEDS network used in the numerical tests

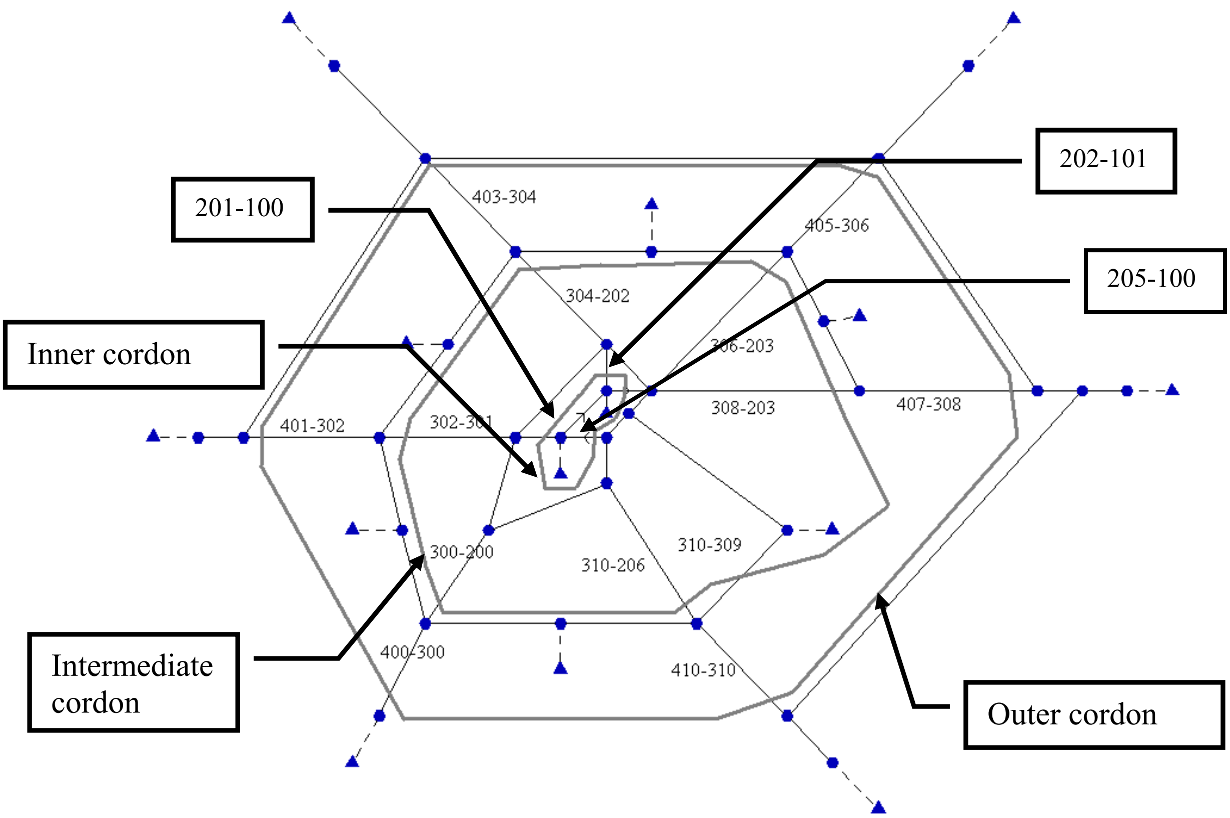


Table 1 shows the GA parameters used in each problem. The issue of choosing the best parameters for GA is still a very active research topic in the meta-heuristic optimisation area. We will not discuss this issue in this paper. Thus, the parameters used in this experiment are still very judgmental and can be adapted to gain a better performance of GA.

Table 1: GA parameters for the tests with MINILEEDS

Problem	Method	Generation numbers	Population numbers	Probability of crossover	Probability of mutation
OPT1: Inner cordon	GA-CHARGE	50	30	0.15	0.05
OPT1: Intermediate cordon	GA-CHARGE	50	30	0.15	0.05
OPT1: Outer cordon	GA-CHARGE	50	30	0.15	0.05
OPT2: 6 links	GA-LOCATE	30	30	0.15	0.05
OPT2: 10 links	GA-LOCATE	30	30	0.15	0.05
OPT3	GA-LOCATEII	50	30	0.15	0.05

3.2 Numerical results

3.2.1 The results of OPT1

Figure 8 shows the three predefined charging cordons, i.e. inner, intermediate, and outer cordons. The CORDON and GA-CHARGE processes are employed to find the optimal toll on each toll point of these three cordons. Table 1, 2, and 3 show the optimal toll levels and benefits found by CORDON and GA-CHARGE for the inner, intermediate, and outer cordons respectively. The percentage in brackets is the percentage of social welfare improvement in first best condition. The first-best condition is to apply the marginal cost tolls derived from the system-optimum assignment on all links.

In our case, the social welfare improvement for the first-best condition is £5,213 per single AM peak period. The optimal uniform tolls around each cordon are also calculated using a standard univariate optimisation method. The optimal uniform tolls for the inner, intermediate, and outer cordons are £0.21, £0.19, and £1.04 with social welfare improvements of £166, £445, and £923 per single AM peak period respectively.

From the tables note that allowing the tolls to vary around the cordons increases the benefits significantly. Furthermore applying GA-CHARGE gives higher benefits than the solution produced by CORDON in all cases[4]. The benefits increase by 23%, 8%, and 12% for the inner, intermediate, and outer cordons respectively. Figure 9 illustrates the process of GA-CHARGE for the outer cordon which consists of 6 tolled links. The Y-Axis is the fitness value of each chromosome which is the value of social welfare improvement (£ per single AM peak). The X-Axis is the chromosome number. Note that in this test, the population size is 30 with 50 generations.

Table 2: Optimal tolls and benefits for the inner cordon in MINILEEDS network calculated by CORDON and GA-CHARGE

Link	Optimal toll from CORDON (£)	Optimal toll from GA-CHARGE (£)
201-100	0.75	0.68
202-101	1.07	0.90
205-100	0.41	0.24
Social welfare improvement (£ per single AM peak)	305 (5.8%)	375 (7.2%)

[4] Note that the CORDON process did not converge properly for the cordons due to subtle changes in the path sets and so the benefits are not necessarily optimal – hence there is room for improvement which is where GA-CHARGE can obtain extra benefits.

Table 3: Optimal tolls and benefits for the intermediate cordon in MINILEEDS network calculated by CORDON and GA-CHARGE

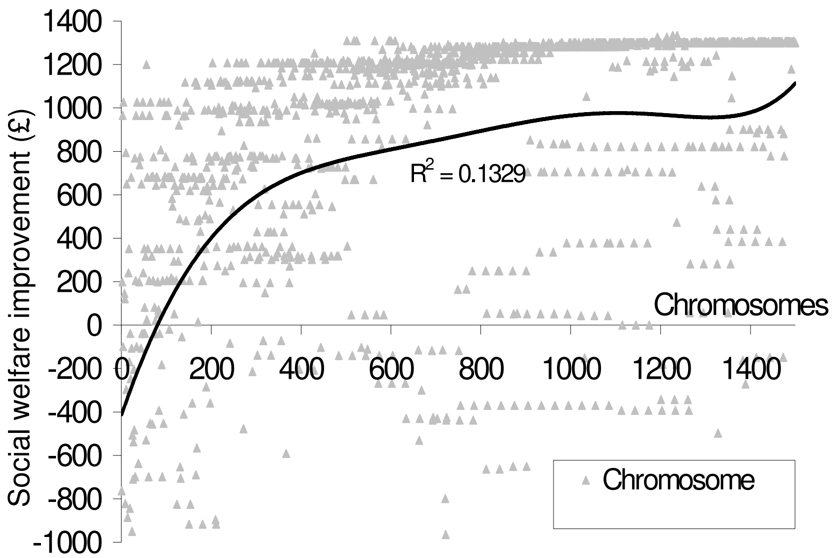
Link	Optimal toll from CORDON (£)	Optimal toll from GA-CHARGE (£)
302-201	0.38	0.34
304-202	0.62	0.63
306-203	0.66	0.73
308-203	0.73	0.73
310-309	0.09	0.09
310-206	0.07	0.08
300-200	0.09	0.09
Social welfare improvement (£ per single AM peak)	1,005 (19.2%)	1,084 (20.8%)

Table 4: Optimal tolls and benefits for the outer cordon in MINILEEDS network calculated by CORDON and GA-CHARGE

Link	Optimal toll from CORDON (£)	Optimal toll from GA-CHARGE (£)
401-302	0.75	0.83
403-304	1.13	1.40
405-306	1.17	2.05
407-308	1.03	1.39
410-310	1.02	1.07
400-300	0.30	0.65
Social welfare improvement (£ per single AM peak)	1,166 (22.3%)	1,305 (25%)

This means GA only sampled 1,500 chromosomes from all possible combinations. According to the chromosome encoding of GA-CHARGE presented earlier, the total number of possible chromosomes is 2^{t^k}. In this case, t is equal to 6 tolled links and k is equal to 10, the number of digits required in binary format to represent the maximum possible toll given as 1000 seconds, resulting in 2^{600} possible combinations. Note the extremely small ratio between the sampled chromosomes and the possible combinations. Figure 9 shows the trend of the fitness value (using regression analysis) and the fitness of all chromosomes as the GA progressed. It can be seen that the trend of the fitness value increased, and finally converged to the solution shown in Table 3.

Figure 9 : The GA-CHARGE process and the trend of improvement in the fitness value

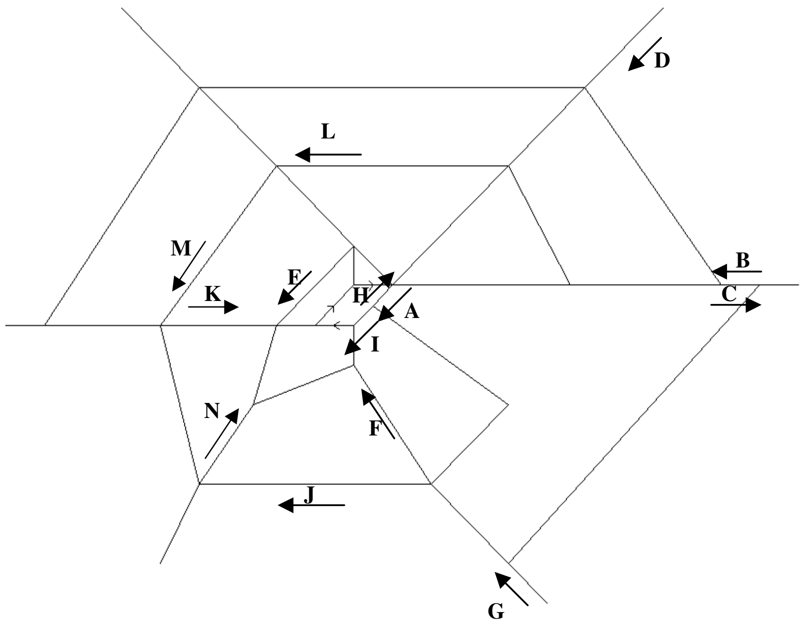


3.2.2 The results of OPT2

The problem of OPT2 is to identify the optimal location of tolled links given the desired number of tolled points. The MINILEEDS network is used again in this test. Two tests are conducted, finding the best six and best ten optimal tolled links. LOCATE and GA-LOCATE are applied to the problems. The total number of directed links in the network is 89. Thus, the possible number of combinations for the problem of six and ten optimal tolled links is approximately 5.8×10^{8} and 5.0×10^{12} respectively, which is impractical to implement through enumeration or greedy search methods. Figure 10 shows the best six and best ten tolled links found by using LOCATE and GA-LOCATE.

Figure 10 is used to explain the results obtained from LOCATE and GA-LOCATE[5]. The arrows in the figure represent the links selected. From Figure 10, LOCATE selected links A, B, C, D, E, and F as the best six tolled links. Then, LOCATE added links G, H, I, and J as the additional four links for the best ten tolled links. On the other hand, GA-LOCATE selected links A, B, C, D, E, and L as the best six tolled links. Note that GA-LOCATE only picked one different link compared to the set of best six tolled links selected by LOCATE (link L rather than link F). GA-LOCATE then selected links A, B, C, D, E, G, I, K, M, and N as the best ten tolled links. Seven links out of ten links selected by LOCATE are also selected by GA-LOCATE.

Figure 10: Location of the best 6 and 10 tolled links from LOCATE and GA-LOCATE



[5] Links B and C are the same link but in the opposite direction. Also, links A and H are the same link but in the opposite direction.

Table 4 shows the social welfare improvement in pounds per single peak hour (the percentage in the bracket is the relative welfare improvement compared to the first-best condition). The optimal benefit of the best six and ten tolled links chosen by GA-LOCATE is only slightly higher than those from LOCATE (approximately 0.9% and 2.9% respectively). The difference is that LOCATE has to include all previously selected links within its solution whereas GA-LOCATE can drop links e.g. link L is not included in the best 10 links even though it is included in the best 6 link solution. It should be noted that even though the LOCATE and GA-LOCATE methods rely on indices which could contain errors similar to those encountered in the CORDON process both methods produce solutions which give rise to 85% of the first best conditions even with only 6 toll points. The fact that adding a further 4 links only gives a marginal increase in benefits suggests that the optimal number of toll points when considering implementation costs would be somewhere between 5 and 10 links.

Table 5: Benefits of the best 6 and 10 tolled links from LOCATE and GA-LOCATE

Method	Benefit for best 6 tolled links	Benefit for best 10 tolled links
LOCATE	£4,385 (84.1%)	£4,611 (88.4%)
GA-LOCATE	£4,427 (86.8%)	£4,745 (91%)

3.2.3 The results of OPT3

In this test, the implementation and operation costs per toll point are calculated using a discounted value over a 30-year period. The cost per toll point is assumed to be £100 per toll point per peak-hour based on estimates by Oscar Faber (2001). GA-LOCATEII is used to find the optimal number of tolled points and their locations. GA-LOCATEII identified 10 as the optimal number of tolled links. It selected the best ten tolled links that were chosen by GA-LOCATE previously. This result is wrong since the net benefit for the best 6 tolled links is £3,827 per single peak hour which is actually higher than that from the net benefit from the best 10 links (£3,745 per single peak hour). The gross location indices for the best 6 and 10 tolled links are £6,465 and £12,192 per single peak hour and the indices net of costs are £5,865 and £11,192. Thus, even after subtracting the costs from the location indices, the set of 10 tolled links remains better than the set of 6 tolled links. The location indices are overestimated for both sets of tolled links. Experience suggests that the toll predictions used in the location index are always an over-estimate of the true optimal tolls i.e. the error terms associated with a link are always positive[6]. Thus we suggest that as the number of links considered is increased then the error in the location index is increased. This does not cause any difficulties for OPT2 as the number of links considered is constant and implementation costs are equal. However as seen here the OPT3 problem has a variable implementation cost and as the magnitude of the errors vary with links considered the solution selects the wrong combination. Further research is required to improve the performance of the location index approach.

5. Conclusions

We have demonstrated that the derivative based approach can solve the second-best tolling problem in most but not all cases. It has been shown to fail due to multiple optima, changes in the path set or as a result of assignment convergence errors. The GA-CHARGE approach was shown to be successful in solving the OPT2 problem giving significant improvements over the CORDON process.

The incremental LOCATE approach performed well in the case study of MINILEEDS, but in general suffers from the weakness whereby previously selected links cannot be de-selected when building a combination of toll points. The GA-LOCATE approach gives only a slight improvement in the case presented as many of the links selected by LOCATE are also in the GA solution. The problem OPT3 is the most difficult problem to solve. The structure of GA-LOCATEII should in theory be able to solve this problem, but errors in the location indices appear to be additive as the number of links considered is increased.

[6] This is our experience with the CORDON process though we have not yet been able to prove this in a general case.

Note that the GA based method is found to be time consuming and there is no proof of convergence of the algorithm. However, the evidence of successful implementation of GA based methods has been growing in the literature. Although our tests are only limited to a simplified network, the results in terms of the relative social welfare improvement values show that GA based methods can find at least a good heuristic optimal solution, particularly for the case of the toll location problem.

Further research will be conducted into improving both the CORDON and LOCATE approaches by using more accurate assignment techniques and including second-order terms. In terms of GA based methods, attention will be paid to the issue of choosing the best parameters for the GA process, i.e. generation number, population number, probability of crossover, and probability of mutation. Finally work is underway to adapt GA-CHARGE to solve OPT3 and include practical cordon design criteria.

References

Abdulaal, M. and LeBlanc, L.J. (1979). "Continuous equilibrium network design models." *Transportation Research* 13B, 19-32.

Allsop, R.E. (1974). "Some possibilities for suing traffic control to influence trip distribution and route choice." In Buckley, D.J. (eds.), *Proc. of the 6th International Symposium on Transportation and Traffic Theory*, Elsevier, New York, 345-373.

Arnott, R.J., de Palma, A. and Lindsey R. (1990). "Departure time and route choice for the morning commuter." *Transportation Research* 42B, 209-228.

Ben-Ayed, O. and Blair, C. (1990). "Computational difficulties of bilevel programming." *Operations Research* 38, 556-560.

Ben-Ayed, O., Boyce, D.E. and Blair III, C.E (1988). "A general bilevel linear programming formulation of the network design problem." *Transportation Research* 22B, 311-318.

Bergendorff, P., Hearn, D.W., and Ramana, M.V. (1997). "Congestion Toll Pricing of Traffic Networks." In Pardalos, P.M., Hearn, D.W., and Hager, W.W. (eds.) *Lecture Notes in Economics and Mathematical Systems Vol. 450*. Springer-Verlag.

Boyce D.E., LeBlanc, L.J., and Jansen, G.R.M. (1991). *Optimal toll stations in a road pricing system: an investigation into the feasibility of optimisation methods*. Report INRO (unpublished).

Cree, N.D., Maher, M.J., and Paechter, B. (1998). "The continuous equilibrium optimal network design problem: a genetic approach." In Bell, M.G.H. (eds.) *Transportation Networks: Recent Methodological Advances*. Pergamon, Amsterdam, 163-174.

Davis, L. and Steenstrup M. "Genetic Algorithms and Simulates Annealing: An overview." in Davis, L. (eds.) *Genetic Algorithms and Simulated Annealing*. Morgan Kaufmann Publishers, Los Altos, California, 1-11.

Friesz, T.L., Cho, H.J., Mehta, N.J., Tobin R.L., and Anadalingam, G. (1992). "A Simulated Annealing Approach to the Network Design Problem with Variational Inequality Constraints." *Transportation Science* 26, 18-26.

Friesz, T.L., Tobin, R.L., Cho, H-J., and Metha, N.J. (1990). "Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints." *Mathematical Programming* 48. 265-284.

Glazer, A. and Niskanen E. (1992). "Parking fees and congestion." *Regional Science and Urban Economics* 22, 123-132.

Hearn, D.W. and Ramana M.V. (1998). "Solving Congestion Toll Pricing Models." In Marcotte, P. and Nguyen, S. (eds.) *Equilibrium and Advances Transportation Modelling*. Kulwer Academic Publishers, Netherlands, 109-124.

Hearn, D.W. and Yildirim, M.B. (2002). "A Toll Pricing Framework for Traffic Assignment Problems with Elastic Demand." In Gendreau M. and Marcotte P. (eds.) *Transportation and Network Analysis: Current Trends*. Kluwer Academic Publishers, Netherlands.

Knight, F.H. (1924). "Some fallacies in the interpretation of social cost." *Quaterly Journal of Economics* 38, 582-606.

LeBlanc, L.J. and Boyce D.E. (1986). "A bilevel programming algorithm for exact solution of the network design problem with user-optimal flows." *Transportation Research* 20B, 259-265.

Levy-Lambert, H. (1968). "Tarification des services a qualite variable: application aux peages de circulation." *Econometrica* 36(3-4), 564-574.

Liu, L.N. and McDonald, J.F. (1999). "Economic efficiency of second-best congestion pricing schemes in urban highway systems." *Transportation Research* 33B(3), 157-188.

Marchand, M. (1968). "A note on optimal tolls in an imperfect environment." *Econometrica* 36, 575-581.

Marcotte, P. (1983). "Network Optimization with continuous control parameters." *Transportation Science* 17, 181-197.

Marcotte, P. (1986). "Network design problem with congestion effects: a case of bilevel programming." *Mathematical Programming* 34, 140-162.

May A.D., Liu R., Shepherd, S.P., and Sumalee A (2002). "The impact of cordon design on the performance of road pricing schemes." *Transport Policy* 9, 209-220.

Meng, Q., Yang, H and Bell, M.G.H. (2001). “An equivalent continuous differentiable model and locally convergent algorithm for the continuous network design problem.” *Transportation Research* 35B, 83-105.

Michalewicz, Z. (1992). *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, New York.

Oscar Faber Consultancy. (2001). *Road User Charging Study-West Midlands, Final Report prepared for Birmingham City Council*. (unpublished)

Pigou, A.C. (1920). *Wealth and Welfare*. Macmillan, London.

Sharp, C.H. (1966). “Congestion and welfare, an examination of the case for congestion tax.” *Economic Journal* 76, 806-817.

Shepherd, S.P., May, A.D. and Milne, D.S. (2001a). “The design of optimal road pricing cordons.” *Proc. of the ninth world conference on transport research*, Seoul, Korea.

Shepherd S.P., May A.D., Milne, D.S. and A. Sumalee (2001b). “Practical Algorithms for Finding the Optimal Road Pricing Location and Charges.” *Proc. of European Transport Conference, September 2001*, Cambridge, UK.

Small, K.A. and Yan, J. (2001). “The Value of “Value Pricing” of Roads: Second-Best Pricing and Product Differentiation.” *Journal of Urban Economics* 49, 310-336.

Smith, M.J. (1979). “The existence, uniqueness and stability of traffic equilibrium.” *Transportation Research* 13B, 295-304.

Steenbrink, P.A. (1974). *Optimization of Transport Network*. John Wiley & Sons, London.

Suwansirikul, C., Friez, T.L., and Tobin, R.L. (1987). “Equilibrium decomposed optimisation: a heuristic for the continuous network equilibrium design problem.” *Transportation Science* 21, 254-263.

Tobin, R.L. and Friesz, T.L. (1983). “Sensitivity analysis for equilibrium network flow.” *Transportation Science* 22(4), 242-250

Van Vliet, D., (1982). “SATURN – a modern assignment model.” *Traffic Engineering and Control* 23(12).

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

Vickrey, W.S. (1969). "Congestion theory and transport investment." *American Economics Review* 59, 251-260.

Walters, A.A. (1961). "The theory and measurement of private and social cost of highway congestion." *Econometrica* 29(4), 676-697.

Wardrop, J. (1952). "Some theoretical aspects of road traffic research." *Proc. of the Institute of Civil Engineers* 1(2).

Wilson, J.D. (1983). "Optimal road capacity in the presence of unpriced congestion." *Journal of Urban Economics* 13, 337-357.

Whitley, D. (1989). "The GENITOR algorithm and selection pressure : Why rank based allocation of reproductive trial is best." In Schaffer, J. (eds.) *Proceeding of the 3rd International Conference on Genetic Algorithms*, Morgan Kaufmann, Publishers, CA, USA.

Yang, H., Zhang, X. and Huang H-J. (2002). "Determination of optimal toll level and toll locations of alternative congestion pricing schemes." In Taylor M.A.P. (eds.) *Transportation and Traffic Theory in the 21st Century: Proceeding of the 15th International Symposium on Transportation and Traffic Theory, Adelaide, Australia,* Elsevier Science, New York.

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

Appendix A: Notation

N	The set of nodes in the network
I	the set of OD-pairs, denoted i=1,…,I
T_i	the continuous number of users (or OD-flow) for OD pair i, with $T_i \geq 0$
$D_i(T_i)$	the inverse demand function for trips for OD-pair i, with $D'_i \leq 0$
J	the set of directed links in the network, denoted j=1,..,J
V_j	the continuous number of users (or link flow) on link j, with $V_j \geq 0$
$C_j(V_j)$	the average cost function for the use of link j, with $c_j' \geq 0$
Cp	the travel costs on path p
$\prod$	the set of non-cyclical paths in the network, denoted p=1,…,P
F_p	the continuous number of users (or path-Flow) for path p, with $F_p \geq 0$
$\prod_I$	the set of non-cyclical paths for OD-pair i, denoted p_i=1,…,P_i
δ_{jp}	A dummy that takes on the value of 1 if link j belongs to path p, and a value of 0 otherwise
ε_j	A dummy that takes on the value of 1 if a toll can be charged on link j, and a value of 0 otherwise
τ_j	the level of the toll on link j if ε_j=1
i or k	index for OD pairs
j or m	index for links
p or q	index for paths
λ_p	Lagrange multiplier associated with path p
Δ_{ip}	A dummy equal to 1 if $p \varepsilon \prod_i$ and