Simultaneous determination of optimal toll locations and toll levels in cordon-based congestion pricing problem  
(case study of Mashhad city)

Sh. Afandizadeh1,*, M. Yadak1, N. Kalantari1  
Received: August 2009, Revised: April 2010, Accepted: June 2010

Abstract
The congestion pricing has been discussed as a practical tool for traffic management on urban transport networks. The traffic congestion is defined as an external diseconomy on the network in transport economics. It has been proposed that the congestion pricing would be used to reduce the traffic on the network. This paper investigates the cordon-based second-best congestion-pricing problems on road networks, including optimal selection of both toll levels and toll locations. A road network is viewed as a directed graph and the cutest concept in graph theory is used to describe the mathematical properties of a toll cordon by examining the incidence matrix of the network. Maximization of social welfare is sought subject to the elastic-demand traffic equilibrium constraint. A mathematical programming model with mixed (integer and continuous) variables is formulated and solved by use of two genetic algorithms for simultaneous determination of the toll levels and cordon location on the networks. The model and algorithm are demonstrated in the road network of Mashhad CBD.

Keywords: Congestion pricing, Genetic algorithm, Network, Cutset, Cordon, Optimization.

1. Introduction
Road pricing has long been recognized as an efficient way to improve the economic efficiency of the transportation system and has been implemented in many metropolises around the world to reduce traffic congestion and pollution. In addition, the revenue from road pricing provides a basis for investment decisions in transportation infrastructure, such as expanding the road capacity, providing better maintenance, and improving public transport. The advanced technology of electronic road pricing mechanisms offers lower cost and new possibilities for road pricing systems. So far, many countries or regions have used pricing systems successfully such as Norway, Singapore, and Hong Kong.

The theory of marginal cost pricing, known as the first-best congestion pricing theory, is well researched and widely advocated by economists. In order to achieve a system optimum flow pattern in the network, a toll equal to the difference between marginal social cost and marginal private cost is charged on each link (Beckmann (1965); Dafermos and Sparrow (1971); Smith (1979)). Special consideration for queues by Yang and Huang (1998) and congestion in a stochastic equilibrium by Yang (1999) has been also investigated. The real world implementation of first best toll pricing has been rare due to public and political resistance and primary due to high extra cost spent on the equipment for toll collection in the entire network.

This has motivated the investigations on second-best pricing, where only a subset of links is subjected to toll charges. In particular, cities such as Singapore, Oslo, Trondheim and Bergen have implemented cordon-based second-best pricing scheme for managing traffic demand.

The problem of two parallel routes where an untolled alternative exists has been investigated for both static and dynamic situations by, for example, Braid (1996), Verhoef et al. (1996), Liu and McDonald (1999), and De Palma and Lindsey (2000), as a special case of second best pricing.
scheme.
Second best congestion pricing in a network is studied by Yang and Lam (1996) for system optimum with fixed demand, by Ferrari (1995) and Yang and Bell (1997), for link flow restriction; to minimize toll revenue subject to a user equilibrium (Hearn and Ramana 1998; Dial 1999a, 2000), and for private highway modeling (Yang and Meng 2000; Yang and Woo 2000).

The second-best pricing has been investigated by Dial (1999b,c); Leurent (1993, 1998) and Yang et al. (2002) for users with discrete or continuous time value distributions.

For the city of Cambridge, in May and Milne (2000) four kinds of road pricing schemes has been compared, with charges based on cordon crossed, distance travelled, time spent in traveling and time spent in congestion, by the SATURN software and its elastic demand assignment response routine SATEASY. A mathematical method to analyze the second-best toll strategies has been proposed by Verhoef (2002) by examining the first-order optimality conditions of the Lagrangian function for social welfare maximization. Various equity issues of the toll design problems were investigated by Yang and Zhang (2002a).

In contrast with the studies on optimal toll pricing, very limited attention has been given to the determination of toll locations on a network. An approach to determine the minimal number of toll links to achieve system optimum was proposed Hearn and Ramana (1998).

Verhoef (2002) examined selection of individual toll links and determination of toll levels using some sensitivity indicators. The performance of various pre-specified toll cordons on a simple hypothetic network has been examined by May et al. (2002a,b). By utilizing the mathematical model proposed by Verhoef (2002), Santos (2002) compared the effect of double cordons and single cordon schemes for seven English towns. Mun et al. (2003) presented a simple spatial model of traffic congestion for a mono-centric city to investigate the effects of cordon pricing on trip-making and congestion level in each location. Without considering the network effect.

Cordon-based congestion pricing, which is already implemented in a number of urban areas, is practically the best way to conduct pricing schemes. Unfortunately, a systematic way to determine various forms of toll cordons and toll levels on a road network is rare in the literature. It is the objective of the present paper to implement models and algorithms for the cordon-based second-best pricing schemes. We consider simultaneous determination of both toll locations and levels for the simplest form of toll cordons, including the single-layered cordon based on Zhang and Yang (2003).

The paper is organized as follows. In the next section, the cutset concept in graph theory is required. For details the reader may refer to Chen (1997). Here we provide a brief introduction to the concept of cutset and its properties.

A graph is a geometrical figure that consists of nodes and links (edges) that connect some of these nodes. Road networks are directed graphs, where the direction of each link represents the direction of vehicle moving. Meanwhile the links possess some physical properties, such as capacity and impedance. In the context of this study, we consider a connected road network on which we can always find at least one path to travel from any node to other nodes (Sheffi, 1985). We also suppose that there is no self-loop link or all links connect different nodes in the network.

The node-link incidence matrix is frequently used in the analysis of graph theory. The node-edge incidence matrix or simply the incidence matrix Aa of an n-node, l-edge graph G without self-loop links is a matrix of order aGl such that

\[
a_{ij} = \begin{cases} 
1 & \text{if link } j \text{ is incident at node } i, \\
0 & \text{if link } j \text{ is not incident at node } i, \\
-1 & \text{if link } j \text{ is incident at node } i, \\
\end{cases}
\]

Thus, in the matrix each row corresponds to a node and each column corresponds to an edge. For convenience, the rank of the incidence matrix of a graph is also termed as the rank of the graph.

A component of a graph is a connected subgraph containing the maximal number of edges. Certainly, a connected graph itself is the biggest component in this graph. The rank of a graph is given by the number of nodes minus the number of components, namely, \( g = a-b \) where \( g \), a and b represent; the rank, number of nodes and number of component in the graph respectively.

A cutset of a graph is defined as a minimal collection of links whose removal reduces the rank of the graph by one (and only one). A single cutset or edge-disjointed union of cutsets of a graph is called a cut. A cut can also be interpreted in another usual fashion. Let V1 be a non-empty proper subset of the node set V of a graph G, and its complement \( V_2 = V - V_1 \). Then the set of links of G each of which is incident with one of its two endpoints in V1 and the other in V2 is a cut of G. In particular, if the removal of these links from G increases the number of components of G by one (and no more than one), then the cut is

2. Preliminary knowledge on graph theory

As mentioned in Zhang and Yang (2003) in order to appropriately choose a toll cordon on a network in a rigorous manner, a preliminary knowledge of the cutset concept in graph theory is required.
also a cutest.

When an area-based pricing scheme is considered for a congested area such as the central business district, the subject area is cordoned off by a simple imaginary closed-loop; the area enclosed within this loop is defined as the cordon area subject to toll charge. It is straightforward to see that the set of links crossing the cordon is a cutset defined in graph theory. Since the cordon divides the whole network into two parts, if we take the nodes in one part as set \( V_1 \) and the other nodes as \( V_2 \), the links crossing the cordon are exactly the set of links with one endpoint in \( V_1 \) and the other in \( V_2 \), which is a cutset by definition. Note that the single-layered cordon divides the whole network into two components; hence the rank of the graph will decrease from to \( n \), where \( n \) is the number of node.

### 3. Optimum design process

In this section a bi-level programming model for determining tolls on a given cordon id presented. The upper-level program is the maximization of social welfare, and the lower-level program is the traffic equilibrium model in terms of the generalized travel cost.

The upper-level social welfare (SW) maximization problem:

\[
\text{max } SW = \sum \omega \left( D_w^{-1}(\omega) - \frac{1}{\beta} \right) \nu_a(x) \quad (1)
\]

Here \( \nu_a(x) \), \( a \in A \) and \( d_w(x), w \in W \) are the solutions of the following traffic equilibrium problem.

The lower-level traffic network equilibrium problem (Sheffi, 1985):

\[
\text{min } \sum \omega \left( C_a(\omega, x) - d_w(x) \right) \quad (2)
\]

Subject to

\[
\sum r \in R_w \nu_r = d_w \quad r \in R_w, w \in W \quad (3)
\]

\[
\nu_a = \sum s \in W \delta_s w a \quad a \in A \quad (4)
\]

\[
f_w \geq 0 \quad r \in R_w, w \in W \quad (5)
\]

The following notation is used:

\( A \) the set of links in the network

\( A_k \) the subset of links on the toll cordon with toll level \( X \)

\( R_w \) the set of all routes between O-D pair \( w \in W \)

\( f_w \) the traffic flow on route \( r \in R_w, w \in W \)

\( V_a \) the flow on link \( a \in A \)

\( V \) a vector of all link flows \( (v_a, a \in A) \)

\( t_a(v_a) \) the travel time on link \( a \in A \) as a continuously increasing function of link flow \( v_a \)

\( d_w \) the demand between O-D pair \( w \in W \)

\( d \) a vector of all O-D demands \( (d_w, w \in W) \)

\( C_a(v_a) \) the generalized travel cost on link \( a \in A \)

\( D_w(\mu_w) \) the demand between O-D pair \( w \in W \) as a function of O-D travel cost \( \mu_w \) between that O-D pair.

\( D_w^{-1}(\omega) \) the inverse of the demand function \( \delta_a \) if route \( a \in A \) \( \mu_w \), and 0 otherwise.

Note that \( C_a(v_a, x) = t_a(v_a) + \frac{x_a}{\beta} \) if \( a \in A \), where \( x \) is the users’ value of time, and \( C_a(v_a, x) = t_a(v_a) \) if \( a \in A \), \( a \in A \); link travel time \( t_a(v_a) \) is a strictly increasing and continuous function of its flow; O-D demand \( D_w(\mu_w) \) is a strictly decreasing and continuous function of the O-D travel time \( \mu_w \).

There are a many methods available to solve the bi-level program (1)-(5), such as genetic algorithm (Afandizadeh, Sh., Shahrabi, R., and Taromi, R., (2006) and Afandizadeh, Sh., Shahrabi, R., and Taromi, R., (2006) or simulated annealing (Afandizadeh, Sh., et al.,(2010)). In this study, the genetic algorithm (GA) method is applied. in order to obtain link flow and O-D demand and the resulting value of the upper-level objective function of social welfare, a traffic assignment has to be executed for each trial \( x \) of the GA algorithm. The procedure of the genetic algorithm used for the determination of an optimal toll level on a given cordon is stated as follows:

**Step 1:** Initial population. Randomly generate initial population of toll levels. For this end, we propose the upper and lower toll levels:

\[
X_i^L \leq X_i \leq X_i^U \quad (6)
\]

Then, the chromosome length is given by equation (7)

\[
n_{i+1} = n_i + 1 \quad (7)
\]

**Step 2:** Function evaluation. Apply the bi-level toll optimization model to obtain maximum social welfare for each given toll level.

**Step 3:** Natural selection. Select those tolls with higher social welfare as survivors and discard the rest.

**Step 4:** Crossover. Conduct pairing among survivors and exchange tolling nodes between pairs.

**Step 5:** Mutation. Randomly modify the parameters of some tolls.

**Step 6:** Next generation. Randomly generate new population of toll levels.

**Step 7:** Verification of stopping criterion. If the stopping criterion has not been reached, go to Step 2; otherwise stop.

As both cordon locations and toll levels are being optimized simultaneously in this paper. As it is hard to formulate and solve the network location problem by traditional discrete optimization methods in view of cordon feasibility requirement for the candidate links and the bi-level cordon toll optimization model. Genetic algorithm has been used as a good method, which is based on the process of population evaluation and natural selection (Haupt and Haupt, 1998). Here the populations are referred to as the feasible cordon and their performances are evaluated by the maximal social welfare that could be achieved with appropriate toll charges on them. Thus, actually a hybrid method has been used in witch one GA is
used to evaluate the fitness of the other.

The single-layered cordon divides the whole network into two subareas, namely the tolling area and the non-tolling area; travelers have to pay as they pass through the links in which divide these two areas. The nodes in the tolling area are tolling nodes and those in the non-tolling area non-tolling nodes.

Instead of determining the optimal toll links directly, tolling nodes (tolling area), are determined in order to find the optimal toll cordon. The procedure of searching for a feasible cordon as mentioned in Zhang and Yang (2003) is as follows. From among the candidate nodes, randomly pick up some as tolling nodes. After a set of tolling nodes are chosen, a cut is determined automatically, which are the links with one endpoint in the tolling area and the other in the non-tolling area.

Then the rank of the incidence matrix of the graph after the removal of the cut is calculated mathematically. If the rank of the graph is equal to \( \alpha - 2 \) (\( \alpha \) is the number of nodes in the network), this cut is accepted as a cutest or a feasible cordon; otherwise, reject it and randomly pick another set of nodes until a new feasible cordon is found. For each given cordon, the bi-level toll optimization model in preceding section is applied to obtain the maximal social welfare and the corresponding optimum uniform cordon toll. Through the process of natural selection in the genetic algorithm, the cordon with optimal performance will be chosen eventually.

Now we consider a simple parameter representation of the GA method for our application. The parameter (gene) here denotes the state of each candidate nodes, each candidate node \( n \in N \), where \( N \) is the set of pre-selected candidate nodes, can be either a tolling node or not. We define that

\[
1 \quad \text{if candidate } n \text{ is selected as a tolling node} \\
0 \quad \text{otherwise}
\]

The candidate \( n \) with \( g_n = I \) form the set of tolling nodes \( N^* \). Therefore, we can encode the parameters (genes) with the simplest binary variable, and 1 bit is enough for each gene, i.e. 0-1 integers.

Row (a) represents the number of candidate tolling nodes. Row (b) represents a chromosome whose length is \( l = 15 \). Hence the set of tolling nodes represented by this chromosome is \( N^* = \{1,2,3,4,5,6\} \). Row (c) represents another chromosome, the associated set of tolling nodes is \( N^* = \{2,3,4,5,6\} \). The procedure used in this paper is depicted in figure 1.

The procedure of the genetic algorithm used for determining an optimal single-layered cordon is stated as follows:

**Step 1:** Initial population. Randomly generate initial population of feasible single-layered toll cordons using the above-mentioned method.

**Step 2:** Function evaluation. Apply the bi-level toll optimization model to obtain maximum social welfare for each given cordon. For this, first apply the genetic algorithm of toll levels for determining the optimal toll on a cordon.

**Step 3:** Natural selection. Select those cordons with higher social welfare as survivors and discard the rest.

**Step 4:** Crossover. Conduct pairing among survivors and exchange tolling nodes between pairs.

**Step 5:** Mutation. Randomly modify the parameters of some nodes.

**Step 6:** Next generation. For each new set of tolling nodes, determine the corresponding cut automatically; maintain those cuts that are cutest (i.e. feasible cordons), and discard the rest; randomly generate new cordons to complement the population.

**Step 7:** Verification of stopping criterion. If the stopping criterion has not been reached, go to Step 2; otherwise stop.

4. Case study of Mashhad

Figure 2 depicts the road network of Mashhad, a city in Iran. The network consists of 141 zones, 1298 nodes, 1726 links.

We applied the methodology proposed in this paper for the network of Mashhad CBD consisting of in total 423 links and 182 nodes. This network is illustrated in figure 3. Homogeneous users are considered. The following BPR function is used to indicate the performance of road links in the network:

\[
t = t_0 \left[ 1 + 0.15 \left( \frac{V}{Q} - 1 \right)^4 \right]
\]

(6)

Where \( d_w \) and \( Q \) are free-flow travel time and capacity of road links respectively. Elastic demand function is used here to describe the reaction of travelers to the generalized travel cost (inclination of time and monetary costs), and the demand function takes the following form for each O-D pair:

\[
d_w = D_w^0 \exp[\rho(-\frac{\mu_w}{\mu_w^0})] \quad w \in W
\]

(7)

Here \( d_w \) is the realized O-D demand, \( D_w^0 \) is the potential O-D demand, \( \mu_w \) is the present O-D travel cost, and \( \mu_w^0 \) is the free-flow O-D travel cost between O-D pair \( w \in W \). The demand elasticity with respect to the O-D travel cost is given by \( \frac{\partial d_w}{\partial \mu_w} \) where \( p \) is regarded as a dimensionless demand elasticity parameter. In this example, the elasticity parameter is taken to be 0.25. The original social welfare in the absence of toll charge is 1613635.

The genetic algorithm described in Section 3.2 is used to determine the optimal single cordon location. In Fig. 3, the nodes from 1 to 50 are set to be candidate tolling nodes. By executing the genetic algorithm, the cordon shown in figure 4 has been eventually selected to be the optimal toll cordon. On this cordon the optimal toll level is 9600 rials, and the corresponding maximum social welfare is 1742692, which represents a welfare increase of 7.99% compared with the non-tolling equilibrium case.

\[
\theta = \frac{1742692 - 1613635}{1613635} \times 100 = 7.99\%
\]

a) Sensitivity analysis of \( p_m \) and \( p_e \).

Now we conduct sensitivity analysis of the \( p_m \) (mutation rate)
and $p_c$ (crossover rate) by changing their values in the genetic algorithms. Table 1 displays the change of the social welfare as parameter $p_m$ varies from 0 to 0.1 and parameter $p_c$ varies from 0.5 to 1.

As shown in table 1 and in figure 5, the maximal social welfare is achieved when $p_c=0.5$ and $p_m=0.06$.

b) Sensitivity analysis of value of time

Figure 6 shows the change of the social welfare as value of time varies from 500 to 3000 tomans per hour. Clearly, the maximal social welfare is achieved when the value of time is equal to 1200 tomans per hour. And shows rapid changes in the range of 500 to 1500 and then has a monotone increase there after.
c) Sensitivity analysis of genetic algorithm iterations
Here, the number of genetic algorithm iterations is used as the stopping criteria.
As shown in figure 7 when the iterations of genetic algorithm are equal to 100, the maximal social welfare is gained.

5. Conclusion

In this paper, we have developed a methodology to determine the toll levels and toll locations simultaneously for the cordon-based pricing schemes. The road network is viewed as a directed graph and the concepts of cutest in graph theory are used to describe the mathematical properties of a toll cordon by examining the incidence matrix of the network. The node-link incidence matrix of a graph is applied to examine its rank, which can identify whether or not a given subset of candidate tolling nodes forms a feasible single-layered cordon. We have used two genetic algorithms to solve the problem of finding the toll levels and toll locations simultaneously. The first genetic algorithm is employed to naturally select the optimal toll cordon by presetting appropriate candidate tolling nodes and examining the resulting node-link incidence matrix.

We have applied the second genetic algorithm to...
determine the optimal toll levels for each given cordon by solving the bi-level optimization model. The proposed methodology is investigated in our case study with Mashhad CBD urban road network. From our numerical experiment results, the cordon based pricing scheme brings 7.99% social welfare improvement. The results have shown that although the optimal cordon has been achieved but this optimal cordon may not be practical and might need major modifications. In the future researches the authors will devise a methodology to overcome this deficiency.

References

Urban Economics 32, 281-310.


