Abstract: Consider a city of an arbitrary shape where difference classes of users are distributed continuously over the city region. Within this region, the road network is dense and can be represented as a continuum and users patronize a two-dimensional continuum transportation system to the central business district. In this paper, two different congestion-pricing models for this continuum transportation system with multiple user classes are studied. The first model is concerned about the social welfare maximization, which determines the optimal toll rates that maximize the total benefit of the whole system, and the second model is cordon-based congestion-pricing, which offers a sub-optimal but more practical tolling strategy. Both of these models are solved by the finite element method and a promising Newtonian-based solution algorithm. Finally, a numerical example is adopted for giving an idea on the impact of cordon toll charges on the resultant social welfare.

Key Words: Multiple user classes, congestion pricing, continuum model, anonymous toll

1. INTRODUCTION

In the literature, the modeling of traffic equilibrium problems is classified into two general approaches: the discrete modeling approach and the continuum modeling approach. The discrete modeling approach, in which each road link within the network is modeled separately and the demand is assumed to be concentrated at hypothetical zone centroids, is commonly adopted for detailed planning. The continuum modeling approach, in contrast, is used for the initial phase of planning and modeling in broad-scale regional studies, in which the focus is on the general trend and pattern of the distribution and travel choice of users at the macroscopic, rather than at the detailed, level. In the continuum approach, the dense network is approximated as a continuum in which users are free to choose their routes in a two-dimensional space. The fundamental assumption is that the differences in modeling characteristics, such as the travel cost and the demand pattern, between adjacent areas within a network are relatively small as compared to the variation over the entire network. Hence, the characteristics of a network, such as the flow intensity, demand, and travel cost, can be
represented by smooth mathematical functions (Vaughan, 1987). A promising extension to modeling the study area of an arbitrary city shape and the advancement of a solution algorithm have recently been made (Wong et al., 1998; Ho et al., 2006), which adopts the finite element method (FEM) to solve the resultant continuum model (Zienkiewicz and Taylor, 1989).

The continuum modeling approach has various advantages over the discrete modeling approach in macroscopic studies with very dense transportation systems (Blumenfeld, 1977; Taguchi and Iri, 1982; Sasaki et al., 1990; Gwinner, 1998). First, the continuum modeling approach is computationally attractive for very dense transportation networks, because problem size in the continuum model depends on the method that is adopted to approximate the modeling region but not on the actual network itself. Therefore, the use of any effective approximation method, such as the finite element method, can extensively reduce the size of the problem. Second, less data is required in setting up a continuum model. As a continuum model can be characterized by a small number of spatial variables, it can be set up with a much smaller amount of data compared to the discrete modeling approach, which requires data for all of the included links. This makes the continuum model suitable for macroscopic studies in the initial phase of design, as the resources for the collection of data in this phase is very limited. Finally, the continuum modeling approach can help to give a better understanding of the global characteristics of a road network. As the numerical results of a continuum model can readily be visualized in a two-dimensional sense, the influence of different model parameters and the spatial interaction between locations can easily be detected and analyzed.

Congestion-pricing can be used to reduce traffic congestion and raise revenue for funding transportation improvements. The underlying principle of congestion-pricing is to regulate the choice behavior of users to approach a system-optimal travel pattern, which maximizes the total benefit of the system by imposing tolls to the users. This congestion-pricing technique has been adopted in both of the discrete and continuum modeling approaches that are discussed in the previous paragraphs. In the discrete modeling approach these congestion charges are applied to the links within the transportation networks (Beckmann, 1965), whereas in the continuum modeling approach, the tolls are charged based on the distance traveled, or more practically, are applied when traveling across the charging cordons as in Ho et al. (2005) in which only a single class of users was considered. However, in a transportation system, users can be classified into different classes by trip purpose, vehicle type, income level, etc. Each user class may have their own criteria in making trips, selecting routes, and may be charged differently. Thus, it is reasonable to incorporate the concept of multiple user classes in congestion-pricing problems, in which a more realistic estimation of the flow pattern and a more efficient charging mechanism can be deduced. In Yang and Huang (2004) a congestion-pricing problem with multiple user classes for the discrete modeling approach was introduced. In their study, the characteristics of multiple user classes were modeled by considering different values of time and an anonymous tolling scheme was adopted on each road link within the network. By solving the dual formulation of the minimization problem, the optimal flow and toll were found for minimizing the total cost of the system.

In this study, we will extend the congestion-pricing model of the continuum modeling approach that was introduced in Ho et al. (2005) to the case of multiple user classes. Two different congestion-pricing models – social welfare maximization model which determines the optimal toll rates that maximize the total benefit of the whole system, and cordon-based
charging model which evaluates the likely benefits of charging road users who pass through a toll-charging cordon – will be investigated and compared in this multi-class congestion-pricing problem. In Section 2, the definitions and notation adopted for this multi-class congestion-pricing problem are introduced. The formulations and solution algorithms of the social welfare maximization model and cordon-based congestion-pricing model are given in Sections 3 and 4. Section 5 presents a numerical example to illustrate the effectiveness of the proposed methodology, and to compare and contrast the results of the two proposed models.

2. DEFINITIONS AND NOTATION

Consider a city with a CBD as shown in Figure 1, in which the road network is approximated as a continuum (Sasaki et al., 1990). It is assumed that the CBD is sufficiently compact, compared with the whole city region. Different classes of users will travel from their demand location in the city region to the CBD. Denote the city region as $\Omega$, the boundary of the city as $\Gamma$, and the location of the CBD as $O$ which is embraced by the CBD cordon $\Gamma_c$ with $\Omega \subset \Gamma_c$. The demand distribution of class $m$ users is represented by $q_m(x,y)$, where $q_m(x,y)dx\,dy$ is the total demand of class $m$ users that is generated from a demand location $(x,y)\in\Omega$ to the CBD. This demand is directly related to the total travel cost of this class of users from their demand location to the CBD, and can be specified as,

$$ q_m(x,y) = D_m(x,y,u_m(x,y)), \quad (1) $$

where $u_m(x,y)$ is the total travel cost of class $m$ users from a demand location $(x,y)\in\Omega$ to the CBD. The demand function $D_m$, which measures the elasticity of the demand with respect to the total travel cost to the CBD, is assumed to be a monotonically decreasing function in $u_m(x,y)$. As $D_m$ is a monotone function, its inverse $D_m^{-1}$ exists.

The unit transportation cost in the city region is assumed to be dependent on the local flow intensity and road configuration, but not upon its direction (i.e. the isotopic case),

$$ c_m(x,y) = a_m(x,y) + b_m(x,y)\sum_n |f_m(x,y)|, \quad (2) $$

where $c_m(x,y)$ is the unit transportation cost for class $m$ users at location $(x,y)\in\Omega$, $a_m(x,y)$ and $b_m(x,y)$ are strictly positive scalar functions of the cost-flow relationship of the class $m$ users that reflects the local characteristics of the road streets, $f_m(x,y) = (f_{mx}(x,y), f_{my}(x,y))$ is a

![Figure 1. The modeled city](image-url)
vector representing the flow state of class \( m \) users in the city, and \( f_{mx}(x,y) \) and \( f_{my}(x,y) \) are the flow flux in the directions \( x \) and \( y \) respectively, and
\[
|\mathbf{f}_m(x,y)| = \sqrt{f_{mx}(x,y)^2 + f_{my}(x,y)^2}
\]  
(3)
is the norm of the flow vector that measures the traffic flow intensity of class \( m \) users at \((x,y)\). In the city region \( \Omega \), the flow vector and trip demand must satisfy the flow conservation condition as given below:
\[
\nabla \cdot \mathbf{f}_m(x,y) - q_m(x,y) = 0, \quad \forall (x,y) \in \Omega, \quad m \in M.
\]  
(4)
where \( M \) is the number of user classes considered in this model. Assuming that there is no traffic flow on and across the boundary of the city, we have
\[
\mathbf{f}_m = 0, \quad \forall (x,y) \in \Gamma, \quad m \in M.
\]  
(5)
However, it is not too difficult to extend the model to deal with the case that \( \mathbf{f}_m \cdot \mathbf{n} = g_m(x,y) \) on the city boundary \( \Gamma \), where \( \mathbf{n} \) is the unit normal vector that points away from the city region on the boundary \( \Gamma \), and \( g_m \) is a function that represents the given demand distribution of class \( m \) users entering or leaving the city through the boundary. From the flow conservation principle, the total flow of class \( m \) users that enters the CBD must be equal to the total demand of that class of users generated from the whole city region. Therefore,
\[
Q_m = \int_{\Gamma_c} \mathbf{f}_m \cdot \mathbf{n_c} \, d\Gamma = \int_{\Omega} q_m(x,y) \, d\Omega,
\]  
(6)
where \( Q_m \) is the total demand of class \( m \) users that is generated in the city region, and \( \mathbf{n_c} \) is the unit normal vector on the CBD cordon \( \Gamma_c \). In this study the cost incurred by class \( m \) users inside the CBD, \( C_m \), is assumed to be a constant. However, it is not too difficult to extend the model to deal with the case that \( Q_m = S_m(Q) \), which depends on the total flow \( Q = \sum Q_m \) that enters the CBD. The total social cost is the sum of the social cost inside the CBD and that in the city region and is defined as:
\[
C_S = \sum_m C_m Q_m + \sum_m \int_{\Omega} c_m |\mathbf{f}_m| \, d\Omega.
\]  
(7)
From the demand function that is defined in (1), the total user benefit can be determined by
\[
B_U = \sum_m \int_{\Omega} \int_{\xi_0}^{\xi_m} D_m^{-1}(\xi) \, d\xi \, d\Omega.
\]  
(8)
Therefore, the net economic benefit, which is a measure of the social welfare, can be obtained by subtracting the total social cost from the total user benefit,
\[
B_E(\mathbf{f}) = \sum_m \int_{\Omega} \int_{\xi_0}^{\xi_m} D_m^{-1}(\xi) \, d\xi - c_m |\mathbf{f}_m| \, d\Omega - \sum_m C_m Q_m.
\]  
(9)

3 SOCIAL WELFARE MAXIMIZATION MODEL (FIRST-BEST SOLUTION)

3.1 Model formulation

By maximizing the net economic benefit expressed in (9), subject to the constraints of flow conservation and boundary conditions, the system optimal solution can be obtained. The
optimization results provide a first-best solution that maximizes social welfare in the city, while taking into account the congestion externalities in the city region. The problem can be formulated as the following mathematical program:

Maximize $B_{\text{c}}(\mathbf{f}) = \int_{\Omega} \sum_{m} \int_{0}^{q_{m}} D_{m}^{-1}(\xi) \, d\xi - \sum_{m} c_{m} |\mathbf{f}_{m}| \, d\Omega - \sum_{m} C_{m} Q_{m}$, \hspace{1cm} (10a)

subject to

$\nabla \cdot \mathbf{f}_{m} - q_{m} = 0, \quad \forall (x, y) \in \Omega, \ m \in M$ , \hspace{1cm} (10b)

$\mathbf{f}_{m} = 0, \quad \forall (x, y) \in \Gamma, \ m \in M$ , \hspace{1cm} (10c)

$\int_{\Gamma_c} \mathbf{f}_{m} \cdot \hat{\mathbf{n}} \, d\Gamma - Q_{m} = 0, \quad \forall (x, y) \in \Gamma_c, \ m \in M$ , \hspace{1cm} (10d)

Consider the following Lagrangian for the above maximization problem,

$\Pi = \int_{\Omega} \sum_{m} \int_{0}^{q_{m}} D_{m}^{-1}(\xi) \, d\xi - \sum_{m} \left( a_{m} + b_{m} \sum_{n} |\mathbf{f}_{n}| \right) |\mathbf{f}_{n}| + \sum_{m} u_{m} (\nabla \cdot \mathbf{f}_{m} - q_{m}) \, d\Omega - \sum_{m} C_{m} Q_{m} + \sum_{m} \int_{\Gamma_c} \mathbf{w}_{m} \cdot \mathbf{f}_{m} \, d\Gamma + \sum_{m} \pi_{m} \left( \int_{\Gamma_c} \mathbf{n} \cdot \hat{\mathbf{n}} \, d\Gamma - Q_{m} \right) \delta_{\pi}$

From the variational principles, we can show that

$\delta \Pi = \int_{\Omega} \sum_{m} D_{m}^{-1}(q_{m}) \delta q_{m} - \sum_{m} \left( b_{m} |\mathbf{f}_{m}| \sum_{n} \frac{\delta \mathbf{f}_{n}}{|\mathbf{f}_{n}|} \right) - \sum_{m} \left( a_{m} + b_{m} \sum_{n} |\mathbf{f}_{n}| \right) \frac{\delta \mathbf{f}_{m}}{|\mathbf{f}_{m}|} + \sum_{m} u_{m} (\nabla \cdot \delta \mathbf{f}_{m} - \delta q_{m}) + \sum_{m} (\nabla \cdot \mathbf{f}_{m} - q_{m}) \delta u_{m} \, d\Omega - \sum_{m} C_{m} \delta Q_{m} + \sum_{m} \int_{\Gamma} \mathbf{w}_{m} \cdot \delta \mathbf{f}_{m} + \mathbf{f}_{m} \cdot \delta \mathbf{w}_{m} \, d\Gamma + \sum_{m} \pi_{m} \left( \int_{\Gamma_c} \mathbf{n} \cdot \delta \mathbf{f}_{m} \, d\Gamma - \delta Q_{m} \right) + \sum_{m} \left( \int_{\Gamma_c} \mathbf{n} \cdot \mathbf{f}_{m} \, d\Gamma - Q_{m} \right) \delta_{\pi}$

From Green’s Theorem and the fact that $u_{m} \nabla \cdot \delta \mathbf{f}_{m} = \nabla \cdot \left( u_{m} \delta \mathbf{f}_{m} \right) - \delta \mathbf{f}_{m} \cdot \nabla u_{m}$, after rearranging we have:

$\delta \Pi = \int_{\Omega} \sum_{m} \left[ D_{m}^{-1}(q_{m}) - u_{m} \right] \delta q_{m} - \sum_{m} \left( \left( a_{m} + b_{m} \sum_{n} |\mathbf{f}_{n}| \right) \frac{\delta \mathbf{f}_{m}}{|\mathbf{f}_{m}|} + \nabla u_{m} \right] \delta \mathbf{f}_{m} + \sum_{m} (\nabla \cdot \mathbf{f}_{m} - q_{m}) \delta u_{m} \, d\Omega - \sum_{m} C_{m} \delta Q_{m} + \sum_{m} \int_{\Gamma} \left( \pi_{m} + u_{m} \right) \mathbf{n} \cdot \delta \mathbf{f}_{m} + \mathbf{f}_{m} \cdot \delta \mathbf{w}_{m} \, d\Gamma + \sum_{m} \int_{\Gamma_c} \left( \mathbf{n} \cdot \mathbf{f}_{m} \right) \, d\Gamma - Q_{m} \right) \delta_{\pi}$

As $\delta \mathbf{f}_{m}$ is an arbitrary function in $\Omega \cup \Gamma_c$, $\delta q_{m}$ and $\delta u_{m}$ are arbitrary functions in $\Omega$, $\delta \mathbf{w}_{m}$ is an arbitrary function in $\Gamma$, $\delta \pi$ and $\delta Q$ are arbitrary scalar functions, and $\delta \mathbf{f}_{m}$ vanishes in $\Gamma$. We can easily show that for the stationary point of the Lagrangian $\delta \Pi = 0$, it follows that

$a_{m} + b_{m} \sum_{n} |\mathbf{f}_{n}| + \sum_{n} b_{n} |\mathbf{f}_{n}| \frac{\mathbf{f}_{m}}{|\mathbf{f}_{m}|} + \nabla u_{m} = 0, \quad \forall (x, y) \in \Omega, \ m \in M$ , \hspace{1cm} (11)

$D_{m}^{-1}(q_{m}) - u_{m} = 0, \quad \forall (x, y) \in \Omega, \ m \in M$ , \hspace{1cm} (12)

$\nabla \cdot \mathbf{f}_{m} - q_{m} = 0, \quad \forall (x, y) \in \Omega, \ m \in M$ , \hspace{1cm} (13)

$\mathbf{f}_{m} = 0, \quad \forall (x, y) \in \Gamma, \ m \in M$ , \hspace{1cm} (14)

$\int_{\Gamma_c} \mathbf{f}_{m} \cdot \hat{\mathbf{n}} \, d\Gamma - Q_{m} = 0, \quad \forall (x, y) \in \Gamma_c, \ m \in M$ , \hspace{1cm} (15)

$\pi_{m} + u_{m} = 0, \quad \forall (x, y) \in \Gamma_c, \ m \in M$ , \hspace{1cm} (16)

$\pi_{m} + C_{m} = 0$ . \hspace{1cm} (17)
Equations (13–15) are constraints in the mathematical program (10). From equations (16) and (17), we have

$$u_m = C_m, \quad \forall (x, y) \in \Gamma_c, \quad m \in M.$$  

(18)

Hence, for all class m users, $u_m$ is equal to the CBD cost $C_m$ at the CBD boundary $\Gamma_c$. Moreover, from equation (11), we can observe that the flow vector of class $m$ users is directly opposite to the gradient of the corresponding Lagrange multiplier $u_m$, i.e.

$$(-f_m) // \nabla u_m, \quad \text{wherever} \ f_m \neq 0, \quad \forall m \in M,$$

(19)

where ‘//’ means that the two vectors are parallel to each other. Now, let us specify the “imaginary” cost function, $\tilde{c}_m$, in the city region:

$$\tilde{c}_m = a_m + b_m \sum_n |f_n| + \sum_n b_n |f_n|, \quad \forall (x, y) \in \Omega, \quad m \in M.$$  

(20)

For any used path $p$ of class $m$ users from their demand location $H$ in $\Omega$ to the CBD, if we integrate the “imaginary” cost along this path, the total cost incurred by the user can be obtained as

$$\overline{C}_{mp} = C_m + \int_p \left( a_m + b_m \sum_n |f_n| + \sum_n b_n |f_n| \right) \frac{f_m}{|f_m|} \ ds = C_m - \int_p \nabla u_m \cdot ds = C_m - (u_m(O) - u_m(H)) = u_m(H)$$

(21)

using equations (11, 19 and 20), and the facts that $f_m / |f_m|$ is a unit vector that is parallel to $ds$ along the path and $C_m = u(O)$. Therefore, the total “imaginary” cost is independent of the used paths and $\overline{C}_{mp} = u_m(H)$. In contrast, for any unused path $\tilde{p}$ between the demand location $H$ in $\Omega$ of class $m$ user and the CBD, the total cost incurred by the user is

$$\overline{C}_{\tilde{mp}} \geq C_m + \int_p \left( a_m + b_m \sum_n |f_n| + \sum_n b_n |f_n| \right) \frac{f_m}{|f_m|} \ ds = C_m - \int_p \nabla u_m \cdot ds = C_m - (u_m(O) - u_m(H)) = u_m(H)$$

(22)

using equations (11, 19 and 20). Hence, $\overline{C}_{\tilde{mp}} \geq u_m(H)$. The inequality in the above derivation is due to the fact that for some segments along the path $\tilde{p}$ the vectors $f_m / |f_m|$ and $ds$ are not parallel and hence $ds > (f_m / |f_m|) \cdot ds$ for these segments of path $\tilde{p}$. Therefore, for any unused paths the total “imaginary” cost is greater than or equal to that of the used paths. In this way, the model will guarantee that system users will choose his/her route in the city in a user-optimal manner with respect to the “imaginary” cost functions. Moreover, the Lagrangian multiplier $u_m(x, y)$ can be interpreted as the minimum “imaginary” cost (including the CBD cost) of class $m$ users from a location $(x, y) \in \Omega$ to the CBD. It is evident that the “imaginary” cost function consists of two components. Consider,

$$a_m + b_m \sum_n |f_n| = \text{the unit transportation cost experienced or perceived by a class } m \text{ user traveling at } (x, y) \in \Omega;$$

$$\sum_n b_n |f_n| = \text{the user externality, which is the cost per unit distance that a marginal user imposes on others already traveling at } (x, y) \in \Omega.$$
that this is a kind of anonymous externality, as this externality is the same for all classes of users.

This means that, if the cost perceived by the individual user is modified by charging an anonymous toll rate of

\[ \tau = \sum_n b_n \left| f_n \right| \quad \forall (x, y) \in \Omega, \]

per unit distance of travel in the city region, then the flow pattern corresponding to social welfare maximization will be in user optimal. With the advance of the intelligent transportation system, charging this toll rate is now feasible by using global positioning system (GPS) technology, by which every unit of movement at location \((x, y)\) in the city region can be charged by an amount that is equivalent to \(\tau(x, y)\) as given in equation (23).

### 3.2 Solution algorithm

Similar to Wong et al. (1998), the finite element method (FEM) is used for approximating the continuum nature of the social welfare maximization problem stated in equation (10) (Zienkiewicz and Taylor, 1989). By expanding the Lagrangian \((\Pi)\) of this maximization problem with Taylor series and neglecting the higher order term, we have:

\[ \Psi_{k+1} = \Psi_k - H(\Psi_k)^{-1}R(\Psi_k) \]

where \(\Psi_k\) is the solution vector at iteration \(k\); \(R(\Psi_k)\) and \(H(\Psi_k)\) is respectively the residual vector and Hessian matrix of the Lagrangian in iteration \(k\). Based on this iterative equation and the golden section method (Sheffi, 1985) for determining the step size \(\lambda\), the solution procedure for this social welfare maximization problem can be set as follows.

**Solution Procedure**

1. Find an initial solution \(\Psi_0\). Set \(k = 0\).
2. Evaluate \(R(\Psi_k)\) and \(H(\Psi_k)\).
3. If the relative error \(\left| R(\Psi_k) \right| / \left| \Psi_k \right|\) is less than an acceptable error \(\varepsilon\), then terminate, and \(\Psi_k\) is the solution.
4. Otherwise, apply the golden section method to determine the step size \(\lambda^*\), which minimizes the norm of the residual vector \(\left| R(\Psi_k - \lambda H(\Psi_k)^{-1}R(\Psi_k)) \right|\). Then, set \(\Psi_{k+1} = \Psi_k - \lambda H(\Psi_k)^{-1}R(\Psi_k)\).
5. Replace \(\Psi_k\) with \(\Psi_{k+1}\). Set \(k = k + 1\) and go to Step 2.

### 4 CORDON-BASED CONGESTION-PRICING MODEL

The previous section provides a formulation of the congestion-pricing problem that maximizes the social welfare of the system (the first-best solution). However, this first-best charging scheme has not been practiced so far. Technologically, the use of GPS allows the location and routing of each vehicle in the city to be traced and charged differently. However, the administrative cost of introducing and enforcing the first-best scenario may be prohibitively high because each trip has to be charged differently according to the exact route
taken and origin of the journey. It is generally difficult for road users to comprehend this charging system. Moreover, this often gives rise to allegations of the intrusion of privacy – a politically sensitive issue in many societies. Hence, the first-best charging scheme remains a theoretical ideal with limited applicability.

In cities and countries where road pricing has been accepted, either road-based or area-based road pricing is practiced (May et al., 2002). In other words, the congestion toll is collected at one or multiple cordon(s) located at specific site(s) in the city region. The use of short-range radio communication technology and the installation of cordon gantries allow the government to detect vehicles passing through the cordon and the toll charged electronically from prepaid cards without slowing down the traffic (Goh, 2002). The cordon-based charging scheme is easier for road users to understand and does not intrude on their privacy. However, the establishment of cordons imposes constraints on the system and the social welfare maximization solution deviates from the first-best scenario. In other words, “one obtains second best results that trade off in a complex way the deviations from full marginal cost pricing in the different transport markets” (Proost et al., 2002).

Apart from the setting of optimal congestion toll levels, researchers and policy-makers face two additional critical issues when the cordon-based charging scheme is adopted. First, how many cordons should be set up? Generally, the setting up of one simple closed cordon road price, like collecting the CBD toll, is the most convenient both for the administrators and the road users. However, May et al. (2002) have shown that a single closed cordon road pricing can give rise to substantially lower economic benefits when compared to a more complex multiple-cordon design. Thus, there is a need to maintain a balance between the simplicity of the cordon-based congestion pricing scheme and gains in social welfare. Second, where should the cordon(s) be located? Once again, the study of May et al. (2002: 219) demonstrated that the overall “performance of cordon pricing depends critically on the locations chosen for the cordons”. While the levels of road pricing have received much attention from researchers, the issue of the optimal locations of the cordons has been closely examined only recently, using the discrete network approach (Zhang and Yang, 2004). As an analogy to the discrete network approach, this section provides a formulation for the multi-layer cordon-based congestion-pricing model for a continuum transportation system.

4.1 Model Formulation

In order to setup a multi-layer cordon-based congestion-pricing model, we need to determine the total toll that a system user has to pay when traveling from his/her demand location to the CBD boundary. Although the minimum “imaginary” cost, \( u_m(x, y) \), can be obtained by solving mathematical program (10), the total transportation cost and toll are not readily available from the solution. To separate the transportation cost and toll from the function \( u_m(x, y) \), we define a toll function, \( T(x, y) \), which measures the total toll that a user has to pay when traveling from location \((x, y)\) to the CBD boundary. At any location, this toll \( T(x, y) \) should be the same for all classes of users as it is proven in section 3 that all users are charged at an anonymous toll rate \( \tau(x, y) \). Obviously, we have \( \nabla T = \tau^* \left( \frac{\mathbf{f}_m}{|\mathbf{f}_m|} \right) \), where \( \nabla T = \left( \frac{\partial T}{\partial x}, \frac{\partial T}{\partial y} \right) \) is the gradient of the function \( T(x, y) \), \( \mathbf{f}_m \) and \( \tau^* \) are the solutions of the flow vector and anonymous toll rate for the class \( m \) users, respectively, which are determined by solving mathematical program (10), and \( \frac{\mathbf{f}_m}{|\mathbf{f}_m|} \) is the unit vector of the
optimal flow pattern. To solve this partial differential equation, we reformulate it as a least square problem as follows:

\[
\min_{T} \int_{\Omega} \left| \nabla T - \tau^* \left( \frac{f_{m}}{f_{m}^*} \right) \right|^2 \, d\Omega, \quad (25a)
\]

subject to

\[T = 0, \quad \forall (x, y) \in \Gamma_c. \quad (25b)\]

This is a post-analysis of the optimal solution obtained from (10). Based on the solution of total toll \(T(x, y)\) determined from (25), we can also construct \(J\) iso-toll contours, \(\Psi_1, \Psi_2, \ldots, \Psi_J\), in the city region. If two consecutive iso-toll contours of levels \(T_j\) and \(T_{j+1}\) are constructed, then these contours can be considered as the cordons and the area that is embraced by these two contours will form a layer that is charged with an flat anonymous toll of \(\tau_{j+1} = T_{j+1} - T_j\) for any user who enters this layer. Consequently, the set of flat anonymous tolls \(\tau = (\tau_1, \tau_2, \ldots, \tau_J)\) forms the cordon-based charging scheme. The resultant user-optimal solution, subject to this set of flat anonymous tolls, can be obtained by simultaneously solving the following sets of differential equations:

\[
\begin{align*}
\nabla \cdot f_{m} - D_m(u_m, \tau) &= 0, \quad \forall (x, y) \in \Omega, \quad m \in M, \quad (26b) \\
\nabla \cdot f_{m} - D_m(u_m, \tau) &= 0, \quad \forall (x, y) \in \Omega, \quad m \in M, \quad (26b) \\
\n\psi(x, y) &= (x, y), \quad \forall (x, y) \in \Gamma, \quad m \in M, \quad (26d)
\end{align*}
\]

where \(D_m(u_m, \tau)\) is the modified demand function, which takes into account the effect of the total toll that a user has to pay from his/her demand location to the CBD on the intensity of demand. It can be easily proved that the solution of this program satisfies the user optimal conditions (Wong et al., 1998).

### 4.2 Solution Algorithm

Similar to the social welfare maximization model, the FEM is used to approximate the continuous variables in the modeled city. As there is no explicit objective function for this multi-class cordon-based congestion-pricing model, the mixed finite element procedure that was developed by Wong et al. (1998) cannot be directly applied. Thus, we adopt the Galerkin formulation of the weighted residual technique (Cheung et al., 1996; Zienkiewicz and Taylor, 1989). By using the Galerkin formulation, differential equations (26a) and (26b) are transformed into the following equivalent integral expressions:

\[
\begin{align*}
\int_{\Omega} \left[ c_{m}(x, y) \frac{f_{m}(x, y)}{f_{m}^*} \right] + \nabla u_{m}(x, y) \psi(x, y) \, d\Omega &= 0, \quad \forall m \in N_M, \psi(x, y) \quad (27a) \\
\int_{\Omega} \left[ \nabla \cdot f_{m}(x, y) - D_m(u_m, \tau) \right] \psi(x, y) \, d\Omega &= 0, \quad \forall m \in N_M, \psi(x, y) \quad (27b)
\end{align*}
\]

where \(\psi(x, y)\) is the trial (or weight) function in the weighted residual technique. Boundary conditions (26c) and (26d) are enforced by taking a zero weight function (Cheung et al., 1996). In the Galerkin formulation, the local interpolation function of the finite element is used as the trial function. The modeling area is first discretized into a finite element mesh, in which the Galerkin formulation is applied at the element level. The governing equations for all user classes at a particular finite element node \(s\) are given as follows:
where \( \Omega_e \) denotes the domain of the finite element \( e \), \( E_s \) is the set of finite elements that connects node \( s \), \( N_e(x,y) \) is the local interpolation function of the finite element that corresponds to node \( s \), \( \mathbf{r}_{sm} \) is the nodal residual vector for class \( m \) users at node \( s \), which represents the extent to which the governing equations (26a) and (26b) are locally satisfied around node \( s \), and \( \Psi \) is the solution vector of the problem. For the global satisfaction of the governing equations, we require that

\[
\mathbf{R}(\Psi) = \text{Col}(\mathbf{r}_{sm}(\Psi)) = \mathbf{0},
\]

which defines a system of non-linear equations. We apply the Newton-Raphson algorithm with a line search to solve the problem, an iterative equation, which is similar equation (24), could be found

\[
\Psi_{k+1} = \Psi_k - \lambda J(\Psi_k)^{-1} \mathbf{R}(\Psi_k),
\]

where \( J(\Psi_k) \) is the Jacobian matrix of vector \( \mathbf{R}(\Psi_k) \) in iteration \( k \), and \( \lambda \) is the step size, which could also be determined by the golden section method. With this iteration equation, the same solution procedure as that is used for the social welfare maximization model can be applied.

4.3 The Second-best Solution

When setting up a cordon-based congestion-pricing scheme, the location of cordons and their corresponding toll level are two critical issues that have to be addressed. Section 4.1 introduces a method for fixing the location of charging cordons from the first-best solution. However, the foregoing flat anonymous tolls always undercharges users, because they are bounded by the first-best level of toll charges. To fine tune the cordon-based charging levels, given the cordons set by the above procedure remain unchanged, we can derive the second-best congestion-pricing scheme by further optimizing the anonymous toll levels by the following bi-level mathematical program. The upper-level subprogram is to maximize the net economic benefit,

\[
\text{Maximize} \quad B_E(\tau) = \sum_m \int_{\Omega_e} \int_{\Omega_e} \left[ \frac{\tau_m(x,y)}{f_m(x,y)} \right] \left[ \frac{\partial u_m(x,y)}{\partial x} \right] N_e(x,y) \, d\Omega,
\]

subject to

\[
\tau_j > 0, \quad j = 1,2,\ldots,J,
\]

where the quantities, \( \bar{q}_m(\tau) \), \( \bar{c}_m(\tau) \), \( \bar{f}_m(\tau) \), and \( \bar{\Omega}_m(\tau) \), are determined by solving set of differential equations (26) at the lower level, in which the set of flat anonymous tolls \( \tau \) is passed from the upper level to the lower level subprogram. Because in practice only very few cordons are introduced or the number of flat anonymous toll variables are very small, the upper-level subprogram can be easily solved using the pattern search method (Hookes and Jeeves, 1961). Note that the resultant second-best congestion-pricing scheme is optimized in the context of the chosen cordon locations. A better second-best scheme can be obtained by
joint optimization of the cordon locations and anonymous toll levels, which is a useful extension of the present methodology that is worth exploring in the future study.

5. NUMERICAL EXAMPLE

In the following numerical example, we consider a monocentric city, as shown in Figure 1, spanning about 35 km from east to west and 25 km from north to south with a CBD located at the south-western corner of the city. In this modeled city, job opportunities are concentrated in the CBD and hence during the morning peak-hour, the predominant traffic flow is from different parts of the city to the CBD. The elastic demand function is specified as:

Class 1 user: \[ D_1 = 80v_d(x, y)\exp(-0.009u_t). \]

Class 2 user: \[ D_2 = 90v_d(x, y)\exp(-0.011u_t). \]

The demand function is dependent on the total travel cost, \( u_m \), of traveling from the demand location continuously dispersed on the 2-dimensional plane to a common and compact destination at the CBD. \( v_d(x, y) = 1.45 - 0.00005d(x, y) \) is the factor that accounts for the variation of the potential demand and \( d(x, y) \) is the distance of a particular point \((x, y)\) from the CBD. This factor increases when the distance from the CBD decreases, which reveals the fact that area that are closer to the CBD are more densely populated. In a real-life application, the demand surface may vary over space according to the zonal population density and other demographic and socio-economic characteristics of the population specified by the modelers.

The unit transportation cost function is specified as:

Class 1 user: \[ c_1(x, y) = 0.7v_a(x, y) + 0.004v_b(x, y)[|f_1| + |f_2|]. \]

Class 2 user: \[ c_2(x, y) = 0.8v_a(x, y) + 0.005v_b(x, y)[|f_1| + |f_2|]. \]

where \( c_m(x, y) \) is measured in HKD per kilometer, and \( f_1 = (f_{x1}, f_{y1}) \) and \( f_2 = (f_{x2}, f_{y2}) \) are flow vectors of class 1 and class 2 users at location \((x, y)\). \( v_a(x, y) = 1.10 - 0.004d(x, y) \) and \( v_b(x, y) = 1.20 - 0.005d(x, y) \) are the factors that accounts for the variation in the location-dependent parameters of the unit transportation cost function. These factors increase when the distance from the CBD decreases, which reveal the network characteristic that junctions are more closely spaced nearer to the CBD. Hence, the parameters of the unit transportation cost function increase. The cost function is flow-dependent and explicitly takes into account the fact that traffic congestion will lead to a higher travel cost through the opportunity cost of waiting in queues or the payment of a congestion toll. Within the CBD, the CBD cost for class 1 and 2 users is taken as 30 and 40 HKD respectively. Given the above definition of the problem, we proceed to apply the continuum approach in solving the traffic equilibrium problems no-toll, first-best, initial-cordon and second-best scenarios.

First, let us start with the no-toll scenario which is also known as the user optimal model. In this scenario, all users choose their optimal routes over the continuum with the lowest transportation cost (without toll). The resultant traffic flows pattern for class 1 users are displayed in Figure 2. Some associated indicators are calculated and listed in Table 1. For this no-toll scenario, we find that the total vehicular traffic entering the CBD is 37,220 vehicles during the peak-hour. The total social cost is HKD4.81 million and the net economic benefit is about HKD3.84 million. In this scenario, no toll revenue disbursement is possible. Hence, the economic benefit is equal to the difference between total user benefit and total social cost. The resultant traffic flow intensity is displayed in Figure 3. The flow intensity increases rapidly as
traffic approaches the CBD, and thus the total travel cost incurred by the users increases by leaps and bounds. Figure 4 shows the total travel cost in a mapping environment, which clearly illustrates that the total travel cost contours in the proximity of the CBD are much more densely spaced than in the rest of the city.

Table 1 Comparison of results for the no-toll, first-best and initial-cordon scenarios

<table>
<thead>
<tr>
<th></th>
<th>No-toll scenario</th>
<th>First-best scenario</th>
<th>Initial-cordon scenario</th>
<th>Second-best scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBD flow (Class 1 users), $Q_1$ (veh/h)</td>
<td>22,778</td>
<td>16,335</td>
<td>17,687</td>
<td>16,845</td>
</tr>
<tr>
<td>CBD flow (Class 2 users), $Q_2$ (veh/h)</td>
<td>14,442</td>
<td>10,396</td>
<td>11,402</td>
<td>10,855</td>
</tr>
<tr>
<td>CBD flow (Total), $Q$ (veh/h)</td>
<td>37,220</td>
<td>26,731</td>
<td>29,089</td>
<td>27,700</td>
</tr>
<tr>
<td>Maximum toll rate, $\tau_{\text{max}}$ (HKD/km)</td>
<td>-----</td>
<td>47.90</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Maximum paid toll, $T_{\text{max}}$ (HKD)</td>
<td>-----</td>
<td>85.90</td>
<td>80.00</td>
<td>86.00</td>
</tr>
<tr>
<td>Flat anonymous toll for the first cordon layer, $\tau_1$ (HKD)</td>
<td>-----</td>
<td>-----</td>
<td>40.00</td>
<td>49.00</td>
</tr>
<tr>
<td>Flat anonymous toll for the second cordon layer, $\tau_2$ (HKD)</td>
<td>-----</td>
<td>-----</td>
<td>20.00</td>
<td>24.00</td>
</tr>
<tr>
<td>Flat anonymous toll for the third cordon layer, $\tau_3$ (HKD)</td>
<td>-----</td>
<td>-----</td>
<td>20.00</td>
<td>13.00</td>
</tr>
<tr>
<td>Total toll received (HKD)</td>
<td>-----</td>
<td>1,584,789</td>
<td>1,309,155</td>
<td>1,462,614</td>
</tr>
<tr>
<td>Total social cost, $C_S$ (HKD)</td>
<td>4,805,402</td>
<td>2,694,769</td>
<td>3,040,262</td>
<td>2,798,155</td>
</tr>
<tr>
<td>Total user benefit, $B_U$ (HKD)</td>
<td>8,649,225</td>
<td>7,039,679</td>
<td>7,351,132</td>
<td>7,119,267</td>
</tr>
<tr>
<td>Net economic benefit / Social welfare, $B_E$ (HKD)</td>
<td>3,843,823</td>
<td>4,344,910</td>
<td>4,310,870</td>
<td>4,321,113</td>
</tr>
</tbody>
</table>

In the no-toll situation, the traffic congestion is very severe especially as the traffic approaches the CBD. As there is no congestion toll, all road users spend their time waiting in queues, time that could be more fruitfully used for other purposes. In order to minimize the loss of time due to congestion, the idea of congestion pricing should be introduced. In this numerical example, the continuum modeling approach is applied to find the first-best solution of social welfare maximization (with the mathematical program expressed in (10)). With the imposition of congestion tolls, a portion of the congestion externality is internalized by the
road users. Given the city configuration, the congestion toll for all class of users is defined by an anonymous toll (equation (23)). The summary results of the first-best solution are presented in Table 1. Under the first-best scenario, the toll rate, $\tau$, varies continuously over space depending on the location and the traffic volume. Geographically, the anonymous toll rate can be shown by the iso-toll rate contours displayed in Figure 5. As the traffic intensity rises sharply near the CBD, the iso-toll contours are very densely packed. The maximum anonymous toll rate of HKD47.9 per km (Table 1) occurs around the CBD where the traffic volume is the heaviest. For road users departing from any location $(x, y)$ in the city region to the CBD, the total toll paid (equation (25)) can be calculated and shown by the iso-toll contours displayed in Figure 6. Figure 6 illustrates the total toll, $T$, is location-dependent and changes continuously over space. The highest total toll of HKD85.9 (Table 1) is paid by those living farthest away from the CBD. Lastly, the spatial variations of total transportation cost, excluding the toll component, are shown in Figure 7.

In the first-best scenario, road users choose their optimal routes by minimizing the sum of total transportation cost and congestion toll. The resultant total social cost is HKD2.69 million. The total user benefit is HKD7.04 million. In this scenario, the toll received is HKD1.58 million. The toll revenue is transferable and can be used for the improvement of the transportation system in the city region (Small, 1992; Goodwin, 1989). When compared to the no-toll scenario, the net economic benefit (user benefit minus social cost) increases to HKD4.34 million. The resultant CBD usage falls to 26,731 vehicles per hour. In other words,
there is a reduction of traffic volume to the CBD by 28.18%. However, it is important to note that the first-best solution is not a situation of zero congestion. In fact, “it would be impossible to provide sufficient capacity to eliminate congestion. Rather, there is an amount of congestion that is just worth what it costs” (Gillen, 1994: 117).

Based on the first-best solution of this numerical example, the location of the cordons and the corresponding toll levels for this continuum transportation system is defined over the city region. For the sake of simplicity, let us assume that only three tolling cordons are to be set up in this city region. Based on the first-best solution shown in Figure 6, we select the iso-toll contours of HKD40, HKD60 and HKD80 for the purpose of setting up cordons. The selection of these contours will lead to the creation of three toll zones in the city outside the CBD. Users entering each of the three-stepped zones will have to pay (an additional) HKD20 or HKD40. In other words, the total toll paid rises in a step-like manner with the increase in distance from the CBD. Figure 8 shows the situation along the XY transverse in the city region. Based on the foregoing restrictions, a new user optimal solution can be obtained, which forms the initial-cordon scenario.

![Figure 8](image-url)

**Figure 8** Toll paid by users whose origins are located along the cross section XY

The results of this initial-cordon scenario are summarized in Table 1. We show that this scheme, while not socially optimal like the first-best solution, still leads to substantial improvement in social welfare as compared to a no-toll scenario. The number of vehicles entering the CBD during the morning peak-hour falls from 37,220 to 29,089. The reduction in flow intensity is less than the first-best scenario (at 26,731 vehicles per hour). Moreover, the total toll collected drops by about HKD0.28 million (at HKD1.31 million). The net economic benefit is reasonably close to the first-best solution and falls short by HKD34,040 (at HKD4.31 million). Given that we deviate from the first-best solution, there is no reason to suggest that the three cordon charges of the first-best solution will still lead to the maximum social welfare with the foregoing initial-cordon scheme. In other words, the user optimal solution may not be the second-best solution. From Figure 8, it can be seen that the toll paid by the users in this scenario is less than that in the first-best solution. Hence it is reasonable to predict that the second-best solution can be obtained by increasing the cordon toll levels. Hence, the pattern search method introduced by Hookes and Jeeves (1961) was adopted for finding the combination of the cordon toll levels that maximize the social welfare under this initial-cordon scheme. When the net economic benefit is maximized, the charges of the three cordons are HKD49.0, HKD24.0 and HKD13.0, which forms the second-best scenario.

When compared to the initial-cordon scenario, the second-best scenario is associated with a very marginal increase of net economic benefit by only HKD10,243 (at HKD4.32 million). The user benefit slightly decreases from HKD7.35 million to HKD7.12 million, and the total
toll received rises from HKD1.31 million to HKD1.46 million. The total traffic flow to the CBD (27,700 vehicles per hour) is more or less the same as the first-best solution (26,731 vehicles per hour) but is lower than the initial-cordon solution (29,089 vehicles per hour) and the no-toll scenario (37,220 vehicles per hour). Figure 8 shows the variations of the total toll paid by road users at various distances from the CBD along the transverse XY. We show that in some locations, users pay more in the second-best solution than in the first-best solution while in other locations they pay less. When compared with the initial-cordon scenario, the second-best scenario derives the set of optimized tolls that are closer to the continuous toll function of the first-best scenario.

6. CONCLUSIONS

We have proposed the multi-class congestion-pricing models for the continuum transportation system with a single CBD. In this study, two different models, namely the social welfare maximization model and the cordon-based congestion-pricing model have been considered and solved. The social welfare maximization model, which is also known as the first-best solution for this congestion-pricing problem, is formulated as a maximization problem. For this congestion-pricing model, it is also proven that there exists an anonymous toll rate for all classes of user to attain the welfare maximized travel pattern. Based on the first-best solution, the initial-cordon scheme was constructed to locate the cordons and deduce a set of initial flat anonymous tolls. A bi-level programming problem was also formulated and solved by means of a Hookes and Jeeves pattern search algorithm to identify the set of flat anonymous toll levels in the cordons, which leads to the second-best scenario. In all of these problems, the continuum models were solved using the efficient finite element method. Numerical example is adopted to compare and contrast the efficiency of the social welfare maximization model and the cordon-based congestion-pricing model. It is found out that the second-best charging scheme, despite of its simple charging method, attains a net economics benefits that is very close to the first–best scenario.

The use of a continuum approach can help to intuitively identify the level of congestion and the external cost throughout the city, which obviates the need of detailed network modeling work and exhaustive evaluations in the early stage of planning. Moreover, such an intuitive comprehension allows a traffic planner to select with ease one or multiple toll cordon(s) over space and to evaluate the impact of cordon toll charges on the resultant net economic benefits and user benefits, which will provide insightful information for the future formulation of a detailed cordon scheme in a discrete network.

ACKNOWLEDGEMENTS

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU 7126/04E).

REFERENCES


