Pergamon

Transpn. Res.-A. Vol. 28A, No. 3, pp. 197-208, 1994

0965-8564/94 $6.00 + .00

QUEUEING AT A BOTTLENECK WITH SINGLE- AND MULTI-STEP TOLLS

CHEN-HSIU LAIH
Institute of Socio-Economic Planning, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

(*Received* 9 *October* 1992; *in revised form* 3 *June* 1993)

Abstract—We consider toll systems designed to alleviate queueing problems that result from a road bottleneck. Optimal road pricing is capable of eliminating queueing time completely, but has practical difficulties because it requires continuously changeable charges. Because of these difficulties, a step-toll scheme has been considered as an alternative to reduce queueing time. However, previous literatures do not contain an analysis of the multi-step toll and the amount of queueing time removal that can be achieved by varying amounts of step tolls. They may be important for policy makers who are considering the introduction of step pricing under some circumstances. This paper is proposed to solve these problems. We show that at most $n/(n + 1)$ of the total queueing time can be eliminated with the optimal n-step toll. Suboptimal n-step tolls are also derived to achieve an under-$n/(n + 1)$ queueing removal (where $n = 1, 2$ and 3).

1. INTRODUCTION

Because of topographic restraints, such as tunnels and bridges, or artificial restraints, such as lane-decreasing regulation, commuting roads often have a narrow segment. A bottleneck is formed in these situations. A queue will be built up whenever the traffic demand exceeds the capacity of the bottleneck. Queueings cause a wasteful loss of time for commuters and result in an inefficient utilization of the public facility.

Especially during the morning rush hour, because most commuters have almost the same preferred arrival time at work, there are often long and persistent queues at the bottleneck. This problem is becoming worse because of the universal tendency toward increased use of the automobile. In this paper we address the feasibility of using congestion tolls (the step toll system) to ease the problem.

As far as we know, Vickrey (1969) first analyzed the use of pricing to alleviate the problem of a queueing bottleneck. In his model, nontoll equilibrium is illustrated and the optimal variable toll, which eliminates queueings completely, is determined to stagger commuters' departure times. Representative research that extended Vickrey's model includes Hendrickson & Kocur (1981), De Palma *et al.* (1983), Smith (1984), Ben-Akiva et al. (1984), Daganzo (1985), Ben-Akiva, et al. (1986), De Palma and Arnott (1986), Cohen (1987), Braid (1989) and Arnott *et al.* (1987, 1990a, 1990b).

Because the optimal toll, which changes continuously through time, may involve practical difficulties in charging a bottleneck, Arnott *et al.* (1990a) considered the optimal *coarse* (single-step) toll as a substitute to reduce queueings. However, their work does not provide a *flexible* step-pricing that can be adopted for alternatives. This may be important because the effect of queueing removal that will be brought about by their optimal coarse toll probably is not the desired one for either the commuters or authorities. Besides, the goal of queueing removal may sometimes be higher than that which an optimal single-step toll can achieve and consequently a multi-step toll is necessary under this condition. This paper is presented to solve these problems.

The main concern of this paper is the determination of the optimal and suboptimal tolls under the single- and multi-step toll systems to achieve the maximum and alternative (less than the maximum) queueing removals, respectively. All of these are important issues for the practical implementation of step tolls. Also, they are issues that have not been discussed previously.

This paper is organized as follows: section 2 describes the general setting of a bottleneck model and specifies the nontoll equilibrium and optimal time-varying toll by an alternative approach. Section 3 develops the single- and multi-step toll systems, and derives the optimal and suboptimal tolls under these systems. Section 4 concludes the paper.

2. A QUEUEING MODEL OF A ROAD BOTTLENECK

We consider a model of a highway facility with a bottleneck. The highway has one access and one exit. Queueing develops when the capacity of the bottleneck is less than the traffic flow. Route segments before and after the bottleneck have sufficient capacity so that no congestion occurs on them. Queueings that result from stochastic changes in capacity, such as car accidents, are beyond the scope of this paper.

We make the following assumptions: first, there are a fixed number of homogeneous auto-commuters who decide their departure times from home rationally based on the commuting cost minimization principle; second, all auto-commuters have the same preferred arrival time, which is the official starting time; third, commuting demand to the bottleneck is perfectly inelastic, i.e. auto-commuters have no alternative choices but to cross the bottleneck to reach their workplaces; fourth, commuting time other than waiting in the queue is a constant for the commuting decision.
The following notations are used in the model:

N : the number of auto-commuters who need to use the bottleneck;
s : capacity of the bottleneck (measured by traffic flow);
t : departure time from home;
t^* : a fixed official starting time at work for all commuters;
t_q : time when queueing begins to build up;
$t_{q'}$: time when queueing begins to dissipate;
$\tilde{t}$: departure time that will allow one to arrive at work on time after queueing;
T_q : travel time period waiting in the queue;
α : the value penalty per minute of T_q (the shadow value of T_q);
T_e : the time period spent at the workplace before the official starting time;
β : the value penalty per minute of T_e (the shadow value of T_e);
T_L : the period by which the arrival time at work exceeds the official starting time;
γ : the value penalty per minute of T_L (the shadow value of T_L); and
TC : total additional commuting costs (per vehicle) incurred due to queueing.

Under the fourth assumption, we may suppose that a commuter arrives at the bottleneck as soon as he departs from home and arrives at his workplace immediately after passing the bottleneck. Consequently, there appear three possible schedulings of arrival patterns at work:

$$\text{If } (t + T_q) < t^* \text{ (time-early schedule), then } T_e = t^* - (t + T_q) > 0, \text{ and } T_L = 0. \quad (1.1)$$

$$\text{If } (t + T_q) = t^* \text{ (on-time schedule), then } T_e = T_L = 0. \quad (1.2)$$

$$\text{If } (t + T_q) > t^* \text{ (time-late schedule), then } T_e = 0, \text{ and } T_L = (t + T_q) - t^* > 0. \quad (1.3)$$

Figure 1 illustrates the above situations with an example. Obviously commuters face a trade-off of choosing one of the following alternatives when making a commuting decision: *first for (1.1)*, leave home early, take a shorter queue but suffer from early arrival at work; *second for (1.2)*, leave home not too late, suffer from a longer queue but arrive at work on time; *third for (1.3)*, leave home late, take a shorter queue but suffer from late arrival at work.

Now consider an unpriced bottleneck. The extra total cost of a commuting trip due to queueings consists of the following parts: first, the discomfort cost of waiting in the

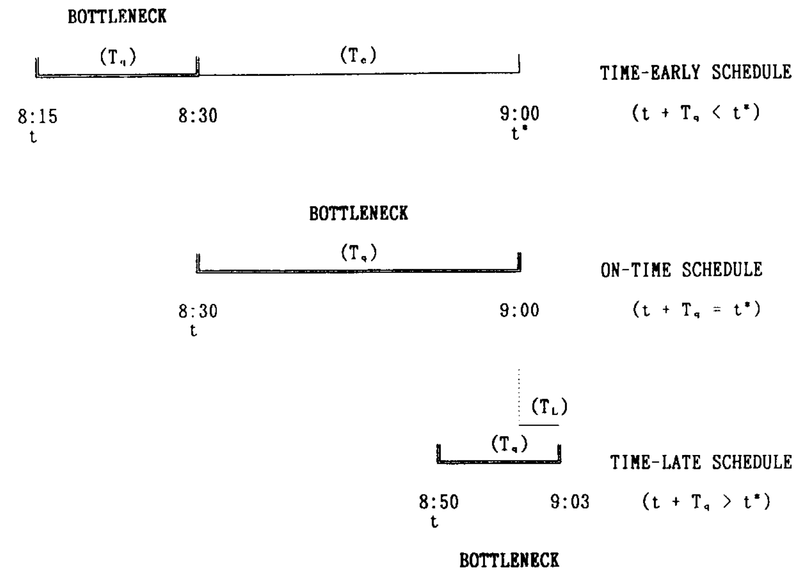


Fig. 1. An example of commuting time schedulings.

queue, and second, the inconvenient (or schedule delay) cost of arriving at work early or late. Thus, the commuter's problem can be expressed as

$$\text{Minimize } TC = \alpha T_q + \beta T_e + \gamma T_L. \tag{2}$$

Here TC is assumed to be linear in T_q, T_e and T_L for simplicity. Substituting (1.1) ~ (1.3) into (2), then we can express the total commuting cost as functions of T_q for the three types of schedulings as

$$TC(t) = \alpha T_q(t) + \beta[t^* - (t + T_q(t))], \quad \text{for } t_q \leqq t + T_q < t^*. \tag{3.1}$$

$$TC(t) = \alpha T_q(t), \quad \text{for } t + T_q = t^* \text{ (or } t = \tilde{t}). \tag{3.2}$$

$$TC(t) = \alpha T_q(t), + \gamma[(t + T_q(t)) - t^*], \text{ for } t^* < t + T_q \leqq t_{q'}. \tag{3.3}$$

Because all commuters seek to minimize total cost, a stable equilibrium will be reached when total commuting costs are equal for all departure times actually used. Therefore, the equilibrium condition is $d(TC)/dt = 0$ for all t. Applying this to (3.1) and (3.3) we obtain:

$$dT_q/dt = \beta/(\alpha - \beta), \quad \text{for } t_q \leqq t + T_q < t^* \tag{4.1}$$

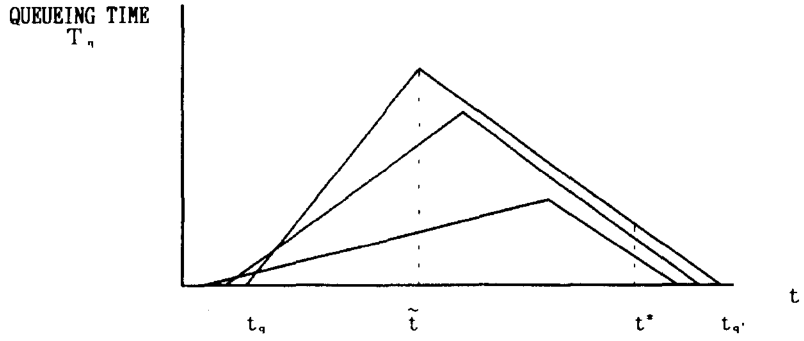
$$dT_q/dt = -\gamma/(\alpha + \gamma), \quad \text{for } t^* < t + T_q \leqq t_{q'}. \tag{4.2}$$

Following Small (1982), we assume that the order relation among time values is

$$\gamma > \alpha > \beta > 0. \tag{5}$$

From (4.1), (4.2) and (5), it is clear that under the nontoll equilibrium, the size of the queue increases linearly from t_q to a maximum at $\tilde{t}$ during the period of time-early schedule, and then decreases linearly from $\tilde{t}$ to $t_{q'}$ during the period of time-late schedule. Also from (4.1) and (4.2), we observe that the shape of queueing time may or may not be symmetric around $\tilde{t}$ depending on whether $\alpha \cdot (\gamma - \beta)$ is equal to $2\beta\gamma$ or not. See three possible cases in Fig. 2.

The decision we now face is where to locate $\tilde{t}$, t_q and $t_{q'}$ under the nontoll equilibrium.



Note: the top triangle is a case of $\alpha(\gamma-\beta) < 2\beta\gamma$

the middle triangle is a case of $\alpha(\gamma-\beta) = 2\beta\gamma$

the bottom triangle is a case of $\alpha(\gamma-\beta) > 2\beta\gamma$

Fig. 2. Possible cases of queueing time under the nontoll equilibrium.

According to Fig. 2 and the definition of $\tilde{t}\,(= t^* - T_q(\tilde{t}))$, we have the following relations:

$$\tilde{t} + [\beta/(\alpha - \beta)](\tilde{t} - t_q) = t^* \text{ or,} \tag{6.1}$$

$$\tilde{t} + [-\gamma/(\alpha + \gamma)](\tilde{t} - t_{q'}) = t^*. \tag{6.2}$$

Next, because the maximum traffic flow that the bottleneck permits is its capacity (s), the total number of commuters (N) who cross the bottleneck during $[t_q, t_{q'}]$ is

$$N = s \cdot (t_{q'} - t_q) \tag{7}$$

Solving (6.1), (6.2) and (7), $\tilde{t}$, t_q and $t_{q'}$ can be determined as

$$\tilde{t} = t^* - [\beta\gamma/\alpha(\beta + \gamma)](N/s). \tag{8.1}$$

$$t_q = t^* - [\gamma/(\beta + \gamma)](N/s). \tag{8.2}$$

$$t_{q'} = t^* + [\beta/(\beta + \gamma)](N/s). \tag{8.3}$$

From (5) and (8.1) ~ (8.3), we have the following order relation:

$$t_{q'} > t^* > \tilde{t} > t_q. \tag{9}$$

Therefore, under the non-toll equilibrium, the official starting time t^* is always within the queueing period $[t_q, t_{q'}]$, so that there will still be queueing during $[t^*, t_{q'}]$ even though the official starting time has passed.

Using (4) and (8), we can obtain the length of queueing time (T_q) for three possible schedulings under the nontoll equilibrium. Because all commuters have the same cost in equilibrium, substituting $T_q(\tilde{t})$ (from (6)) into (3.2) gives the equilibrium commuting cost (TC^e) per auto-commuter as

$$TC^e = [\beta\gamma/(\beta + \gamma)](N/s). \tag{10}$$

Therefore, the nontoll equilibrium can be achieved when all commuters rationally determined their departure times at which their total commuting costs are equal to (10). Namely, no individual has an incentive to change his departure time under this situation.

From (10) we can see that only the inconvenient arrival costs (β and γ) are involved in the equilibrium commuting cost. That is, TC^e is independent of α.

An optimal toll has been defined as a toll that will completely eliminate the efficiency loss due to queueing at a bottleneck (see Vickrey, 1969; Cohen, 1987; Braid, 1989; Arnott, *et al.*, 1990a, etc.). Thus, in order to attain the commuting optimum with the least resistance from the public, we consider an optimal time-varying toll, $\tau(t)$, that will result in T_q being equal to zero and that will leave commuters no worse off than they would be in the nontoll equilibrium. From (3), we obtain such an optimal time-varying toll as

$$\tau(t) = TC^e - \beta(t^* - t) \quad \text{for } t_q \leqq t < t^*. \tag{11.1}$$

$$\tau(t) = TC^e \quad \text{for } t = t^*. \tag{11.2}$$

$$\tau(t) = TC^e - \gamma(t - t^*) \quad \text{for } t^* < t \leqq t_{q'}. \tag{11.3}$$

Figure 3 illustrates the shape of optimal time-varying tolls. Please note that the maximum optimal time-varying toll occurs at the official starting time t^*, as can be observed from (11.2). This is because the penalty costs of β and γ will become zero if one pays the highest optimal toll, $\tau(t^*)$, to arrive at work on time, while (11.1) and (11.3) result in the lower optimal time-varying tolls because they involve penalty costs of β or γ.

Although the nontoll equilibrium and optimal time-varying toll have been derived in previous papers (see Arnott, *et al.*, 1990a, etc.), the proposed approach in this section makes these derivations more simple because the total commuting cost can be simplified as only a linear function of queueing time. This also helps us to develop the derivation of the step toll system in the next section.

A common objection to the optimal toll is that the toll collection is complicated. This is because, for each arriving vehicle, toll collectors must compute variable rates based upon time schedules. Meanwhile, commuters may become impatient if they must spend additional transaction time at the tollgate. To overcome this shortcoming, an Electronic Road Pricing (ERP) system has been experimented with in Hong Kong (1983 ~ 1985). This worked well, and it was estimated that congestion could be reduced by 20% during the morning rush hour (see Dawson and Brown, 1985; Borins, 1988; Hau, 1990 and Laih, 1990). However, there was strong opposition from the public, and the system was finally rejected by Hong Kong's councils.

3. THE STEP TOLL SYSTEM

An optimal time-varying toll can eliminate queueings completely at a road bottleneck. However, there are some financial problems with implementation of such a toll. First, the fixed costs of establishing this pricing system may be high and unacceptable to the authorities. For example, the electronic road pricing system may be expensive. Second, cars must be equipped with an inductive device, such as the electronic number plate, and the expense of renting or buying the device might result in a high fixed cost to

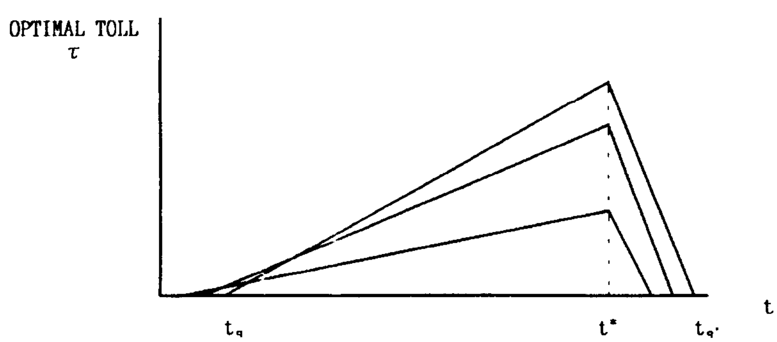


Fig. 3. The optimal time-varying toll for Fig. 2.

commuters. These budget considerations for both the executor and the payer may result in strong resistance to the optimal time-varying toll.

On the other hand, although the step toll system (including the single- and multi-step tolls) can't remove queueings completely, it can be easily implemented by using existing toll gates in the front of bridges or tunnels. Thus, it may be more acceptable than the optimal time-varying toll. Please notice that a uniform toll applied over the queueing period will have no effect on alleviating queueings because the travel demand is inelastic in our model.

Because any tolls that are set to be higher than the optimum time-varying toll (see 11) during a fixed period of time will bring commuters a higher total cost than the nontoll equilibrium cost for the same period of time. In order to avoid such a perverse toll, we assume that the magnitude of the step toll system that we will introduce in the following sections can be no greater than the magnitude of the optimal time-varying toll.

3.1. The single-step toll

We consider a single-step toll, ρ, which will go into effect at time $t^+ \in (t_q, t^*)$, and will go out of effect at time $t^- \in (t^*, t_{q'})$ at the front of the queue. In order to derive the optimal single-step toll, which is defined to maximize the queueing removal (or toll revenue) we design this pricing scheme as $\rho = \tau(t^+) = \tau(t^-)$. Figure 4 illustrates the single-step toll system.

We also suppose that the authorities would like to eliminate $1/\varepsilon$ (where ε is variable and $\varepsilon > 1$) of the total queueing time that exists under the nontoll equilibrium. This is a sensible objective because practical problems may arise if the optimal step-toll with the minimum ε is not the desired amount for either the commuters or authorities under some circumstances.

Our purpose is to find t^+, t^- and ρ for a given ε. In this respect, we have

$$\alpha\left[\frac{\beta\gamma}{2\varepsilon\alpha(\beta+\gamma)}(N^2/s)\right] = (t^- - t^+)\rho\text{'s}. \tag{12.1}$$

$$\rho = \tau(t^+) = \frac{\beta\gamma}{\beta+\gamma}(N/s) - \beta(t^* - t^+). \tag{12.2}$$

$$\rho = \tau(t^-) = \frac{\beta\gamma}{\beta+\gamma}(N/s) - \gamma(t^- - t^*). \tag{12.3}$$

From the previous section, we know that the toll revenue raised by the optimal time-varying toll (the area of any cases of Fig. 3), which will eliminate queueing time completely, is equal to the product of the total queueing time (the area of any cases of Fig. 2: $\{\beta\gamma/[2\alpha(\beta+\gamma)]\}(N^2/s)$) by the queueing costs of α. Consequently, because ρ does not alter trip price, (12.1) states that the revenue of the single-step toll, $(t^- - t^+)$ ρ's, equals $1/\varepsilon$ reduction in the total queueing cost. This equation saves us from solving for the explicit evolution of the queue. Eqns (12.2) and (12.3) result from Fig. 4 directly. Solving these equations yields

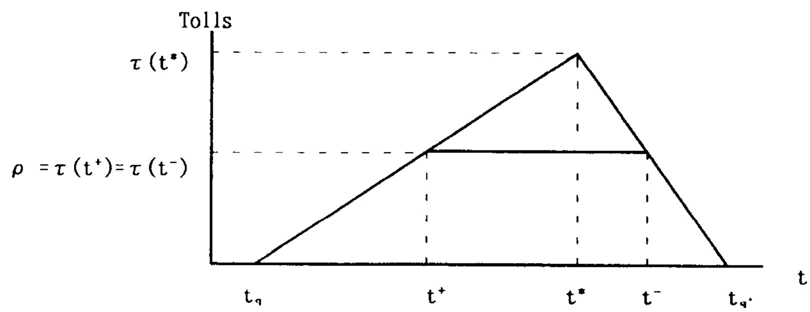


Fig. 4. The single-step toll (ρ) inscribed in the optimal time-varying toll.

$$\rho = \frac{\beta\gamma[1 \pm (1 - 2/\varepsilon)^{1/2}]}{2(\beta + \gamma)} (N/s). \tag{13.1}$$

$$t^+ = t^* - \frac{\gamma[1 \mp (1 - 2/\varepsilon)^{1/2}]}{2(\beta + \gamma)} (N/s). \tag{13.2}$$

$$t^- = t^* + \frac{\beta[1 \mp (1 - 2/\varepsilon)^{1/2}]}{2(\beta + \gamma)} (N/s). \tag{13.3}$$

Because (13) makes sense only when $\varepsilon \geqq 2$, it is clear that the minimum value of ε that maximizes the queueing reduction of (12.1) is equal to 2. Then we obtain the optimal single-step toll in (13) if $\varepsilon = 2$. Thus, we have the following *PROPOSITIONS*:

PROPOSITION 1.1: The optimal single-step toll can eliminate at most half of the total queueing time that exists under the nontoll equilibrium.

Proof: The purpose is to find $\varepsilon\ (= 2)$ that maximizes (12.1).
From (12.2) and (12.3), because $\rho = \tau(t^+) = \tau(t^-)$, we have

$$t^+ = [(\beta + \gamma)/\beta]t^* - (\gamma/\beta)t^-. \tag{P.1-1}$$

Substituting (P.1-1) and (12.3) into the right hand side of (12.1), we obtain

$$(t^- - t^+) \cdot \rho = \{-\gamma(\beta + \gamma)/\beta\}(t^-)^2 + \{[2\gamma(\beta + \gamma)t^* + \beta\gamma(N/s)]/\beta\}t^- - \{[\gamma(\beta + \gamma)(t^*)^2 + \beta\gamma(N/s)t^*]/\beta\}. \tag{P.1-2}$$

Differentiating (P.1-2) with respect to t^- and setting it equal to zero, we have

$$t^- = t^* + \{\beta/[2(\beta + \gamma)]\}(N/s). \tag{P.1-3}$$

Substituting (P.1-3) into (P.1-2) yields

$$\text{Max } (t^- - t^+) \cdot \rho = [\beta\gamma/4(\beta + \gamma)](N/s)^2. \tag{P.1-4}$$

Comparing (P.1-4) with (12.1), we obtain $\varepsilon = 2$. Q. E. D.

PROPOSITION 1.2: The optimal single-step toll is half of the maximum optimal time-varying toll (or nontoll equilibrium cost).

Proof: Comparing (10) and (13.1), we have $\rho = [\beta\gamma/2(\beta + \gamma)](N/s) = \tau(t^*)/2 = TC^e/2$ as given $\varepsilon = 2$. Q. E. D.

On the other hand, there are two sub-optimal single-step tolls in (13) if $\varepsilon > 2$. One is a larger toll with a shorter charging interval, and the other is a smaller toll with a longer charging interval. Because both of them collect the same toll revenue, the authorities face a trade-off of choosing "smaller charge" or "shorter charging interval" to achieve an under-half queueing removal.

3.2. The multi-step toll

Because the single-step toll can only remove up to 50% of the total queueing time that exists under the nontoll equilibrium, we will develop the multi-step toll (also paid at the front of the queue) if the desired reduction is over 50%.

We discuss the double-step tolls first. We design this pricing scheme as the second step toll μ to be higher than the first step toll ρ. Also, similar to the single-step toll, the double-step toll is inscribed in the optimal time-varying toll, $\tau(t)$, so as to derive the optimal and suboptimal solutions to μ and ρ. Therefore, as illustrated in Fig. 5, the lower-charge interval $[t^+, t^-]$ of ρ is discontinuous during the higher-charge interval $[t_+, t_-]$ of μ.

According to Fig. 5, our object is to find t^+, t^-, t_+, t_-, ρ and μ for a given ε (where $1 < \varepsilon < 2$). In this respect, using (12.2) and (12.3) and adding the following equations:

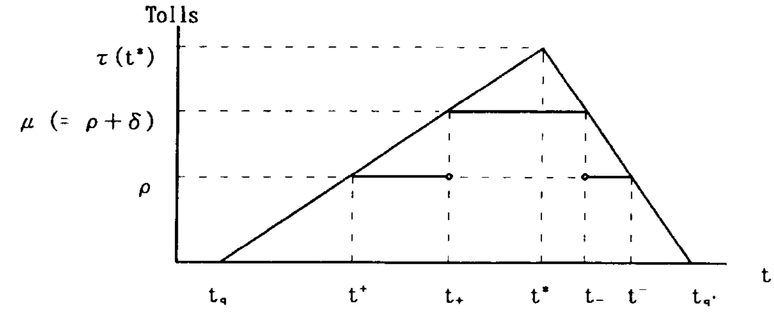


Fig. 5. The double-step toll (ρ, and μ) inscribed in the optimal time-varying toll.

$$\alpha\left[\frac{\beta\gamma}{2\varepsilon\alpha(\beta+\gamma)}(N^2/s)\right] = [(t_+ - t^+ + t^- - t_-)\rho + (t_- - t_+)\mu]s \tag{14.1}$$

$$\mu = \tau(t_+) = \frac{\beta\gamma}{\beta+\gamma}(N/s) - \beta(t^* - t_+) \tag{14.2}$$

$$\mu = \tau(t_-) = \frac{\beta\gamma}{\beta+\gamma}(N/s) - \gamma(t_- - t^*) \tag{14.3}$$

$$\mu = \rho + \delta\ (\delta > 0), \tag{14.4}$$

then we are able to solve for these variables.

Implications of the first three equations are similar to (12). Eqn (14.4) states that the different amount between μ and ρ is δ. δ is assumed to be a positive parameter here because μ is larger than ρ. Solving (12.2), (12.3) and (14) yields

$$\rho = \frac{\beta\gamma(N/s) - \delta(\beta+\gamma)}{2(\beta+\gamma)} \pm \zeta \tag{15.1}$$

$$\mu = \frac{\beta\gamma(N/s) + \delta(\beta+\gamma)}{2(\beta+\gamma)} \pm \zeta \tag{15.2}$$

$$t^+ = t^* + \frac{\rho}{\beta} - \frac{\gamma}{\beta+\gamma}(N/s)\ ;\ t^- = t^* - \frac{\rho}{\gamma} + \frac{\beta}{\beta+\gamma}(N/s) \tag{15.3}$$

$$t_+ = t^* + \frac{\mu}{\beta} - \frac{\gamma}{\beta+\gamma}(N/s)\ ;\ t_- = t^* - \frac{\mu}{\gamma} + \frac{\beta}{\beta+\gamma}(N/s), \tag{15.4}$$

where

$$\zeta = \frac{\beta\gamma}{2(\beta+\gamma)}\left(\frac{2\delta\beta\gamma(\beta+\gamma)(N/s) - 3\delta^2(\beta+\gamma)^2}{\beta^2\gamma^2} - \frac{(2-\varepsilon)(N/s)^2}{\varepsilon}\right)^{1/2}.$$

Because ζ should not be negative, to ε, we have the following constraint:

$$\varepsilon \geqq \frac{2(N/s)^2}{2\delta(\beta+\gamma)(N/s)/(\beta\gamma) - 3\delta^2(\beta+\gamma)^2/(\beta\gamma)^2 + (N/s)^2}. \tag{16.1}$$

We may observe that the function of δ in the denominator of (16.1) has a maximum when

$$\delta^* = \frac{\beta\gamma}{3(\beta+\gamma)}(N/s). \tag{16.2}$$

Thus, substituting (16.2) into (16.1) we obtain the minimum value of $\varepsilon = 3/2$.

Using (16.2) and specifying $\varepsilon = 3/2$, then we have the optimal double-step toll in (15) (where ζ becomes zero). Therefore, we have the following *PROPOSITIONS*:

PROPOSITION 2.1: The optimal double-step tolls can eliminate at most 2/3 of the total queueing time that exists under the nontoll equilibrium.

Proof: The purpose is to find ε (= 3/2) that maximizes (14.1).
Substituting (14.4), (15.3) and (15.4) into the right hand side of (14.1) yields

$$(t_+ - t^+ + t^- - t_-)\rho + (t_- - t_+)\mu$$
$$= [\delta(\beta + \gamma)/\beta\gamma]\rho + [N/s - (\beta + \gamma)(\rho + \delta)/\beta\gamma)][\rho + \delta]. \quad \text{(P.2-1)}$$

Differentiating (P.2-1) with respect to ρ and setting it equal to zero, we have

$$\rho = [\beta\gamma/2(\beta + \gamma)](N/s) - \delta/2. \quad \text{(P.2-2)}$$

Substituting (P.2-2) into (P.2-1), we obtain

$$\text{Max } (t_+ - t^+ + t^- - t_-)\rho + (t_- - t_+)\mu$$
$$= \delta(N/s)/2 - 3\delta^2(\beta + \gamma)/4\beta\gamma + \beta\gamma(N/s)^2/4(\beta + \gamma). \quad \text{(P.2-3)}$$

Because (P.2-3) is a function of δ and has a maximum when

$$\delta = [\beta\gamma/3(\beta + \gamma)](N/s). \quad \text{(P.2-4)}$$

Substituting (P.2-4) into (P.2-3) yields

$$\text{Max } (t_+ - t^+ + t^- - t_-)\rho + (t_- - t_+)\mu = [\beta\gamma/3(\beta + \gamma)](N/s)^2. \quad \text{(P.2-5)}$$

Comparing (P.2-5) with (14.1), we obtain $\varepsilon = 3/2$. Q. E. D.

PROPOSITION 2.2: The optimal double-step toll divides the maximum optimal time-varying toll (or nontoll equilibrium cost) into three equal amounts.

Proof: $\tau(t^*) - \mu = \mu - \rho = \rho = [\beta\gamma/3(\beta + \gamma)](N/s) = \tau(t^*)/3$ as given $\delta = \delta^*$ and $\varepsilon = 3/2$ in (15.1) and (15.2) (where ζ becomes zero). Q. E. D.

If $\varepsilon > 3/2$ on the other hand, then δ in eqn (16.1) can be obtained as

$$\delta^* = \left(\frac{\beta\gamma}{3(\beta + \gamma)} \mp \frac{\beta\gamma[2\varepsilon(2\varepsilon - 3)]^{1/2}}{3\varepsilon(\beta + \gamma)}\right)(N/s). \quad (16.3)$$

Because both solutions of (16.3) exist (i.e. the former in the large parentheses is larger than the latter) if $\varepsilon < 2$, substituting (16.3) into (15) we then have two suboptimal double-step tolls (where ζ becomes zero) if over 1/2 and under 2/3 of the queueing removal is a desirable goal. They are different magnitudes of a smaller toll of ρ and a larger toll of μ. Because both raise the same toll revenue, similar to the case of suboptimal single-step toll, the authorities face a trade-off in choosing one of them at practical requests.

Besides, there is only one solution (the positive one) in (16.3) if $\varepsilon \geqq 2$. This means that there is only one suboptimal double-step toll to remove less than (or equal to) half of the total queueing time. In practice, however, it may be not necessary to consider the double-step toll to remove a fixed amount of queueing time that can be achieved with the single-step toll.

The development to the triple-step toll is similar to the double-step toll. We neglect the repeated derivation of this and provide the results only. According to Fig. 6, we have

$$\rho = \frac{\beta\gamma(N/s) - \eta(\beta + \gamma)}{2(\beta + \gamma)} \pm \xi \quad (17.1)$$

$$\mu = \frac{\beta\gamma(N/s) + (2\delta - \eta)(\beta + \gamma)}{2(\beta + \gamma)} \pm \xi \quad (17.2)$$

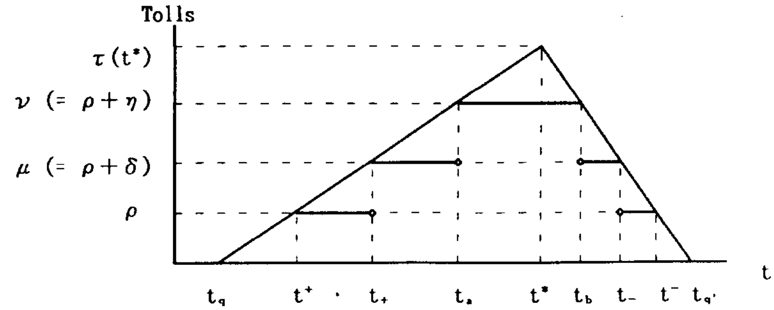


Fig. 6. The triple-step toll (ρ, μ, and ν) inscribed in the optimal time-varying toll.

$$\nu = \frac{\beta\gamma(N/s) + \eta(\beta + \gamma)}{2(\beta + \gamma)} \pm \xi, \tag{17.3}$$

where

$$\xi = \frac{\beta\gamma}{2(\beta + \gamma)} \left(\frac{(\beta + \gamma)^2(-4\delta^2 + 4\delta\eta - 3\eta^2) + 2\eta\beta\gamma(\beta + \gamma)(N/s)}{\beta^2\gamma^2} - \frac{(2 - \varepsilon)(N/s)^2}{\varepsilon} \right)^{1/2}.$$

In Fig. 6, t^+, t^-, t_+ and t_- have the same structures as (15.3) and (15.4). In addition, t_a and t_b are also the same as (15.3) by replacing ρ with ν. ν is the highest toll, charged during $[t_a, t_b]$, of this pricing scheme. The difference between ν and ρ is assumed to be a positive parameter, η, which is larger than the other positive parameter, δ.

Because ξ in (17) should not be negative, we have the following constraint on ε:

$$\varepsilon \geqq \frac{2(N/s)^2}{\{(\beta + \gamma)/\beta\gamma\}\{2\eta(N/s) + [(\beta + \gamma)(-4\delta^2 + 4\delta\eta - 3\eta^2)]/\beta\gamma\} + (N/s)^2}. \tag{18.1}$$

Because the function of δ and η in the denominator of (18.1) has a maximum when

$$\delta^* = \frac{\beta\gamma}{4(\beta + \gamma)} (N/s) \tag{18.2}$$

$$\eta^* = \frac{\beta\gamma}{2(\beta + \gamma)} (N/s), \tag{18.3}$$

substituting (18.2) and (18.3) into (18.1) we obtain the minimum $\varepsilon = 4/3$.

Using (18.2), (18.3) and specifying $\varepsilon = 4/3$, we then have the optimal triple-step tolls in (17) (where ξ becomes zero). Based on these results, we have the following *PROPOSITION*:

PROPOSITION 3: The optimal triple-step toll divides the maximum optimal time-varying toll into four equal amounts and eliminates at most 3/4 of the total queueing time that exists under the nontoll equilibrium.

The proof to the PROPOSITION is similar to PROPOSITIONS 2.1 and 2.2. From (18.1) on the other hand, we have two solutions for δ or η if η or δ is given under the condition of $4/3 < \varepsilon < 3/2$. For example, let $N/s = 2$ hours, $\beta = \$3.9$/hour, $\gamma = \$15.21$/hour (they are chosen on the basis of Arnott, *et al.*, 1990a), $\varepsilon = 1.4$ ($>4/3$) and $\delta = \delta^* = \$1.55$. Substituting these values into eqn (18.1), we obtain $\eta = \$4.06$ or $\$2.14$. Therefore, either δ or η must be given in advance to derive the suboptimal triple-step tolls that can remove more than 2/3 and less than 3/4 of the total queueing time.

In the same way, we can derive the optimal and suboptimal n-step tolls (where $n > 3$). The former divides the maximum optimal time-varying toll into $(n + 1)$ equal amounts and removes up to $n/(n + 1)$ of the total queueing time that exists under the nontoll equilibrium. The latter involves $(n - 1)$ parameters, such as δ and η, which we have assumed for the triple-step toll, and consequently $(n - 2)$ parameters have to be given first to compute the suboptimal n-step tolls if more than $(n - 1)/n$ and less than $n/(n + 1)$ of the total queueing removal is a desirable goal.

It is clear that the step toll system can remove more of the queueing time that exists under the nontoll equilibrium as the number of step becomes more, i.e. $\varepsilon \rightarrow 1$ as $n \rightarrow \infty$. However, too many steps not only confuse commuters in the determinations of their departure times, but also cause the toll collection to become complicated. Because the optimal triple-step toll can make at most only a 8.3% improvement in queueing removal over the optimal double-step toll (the more steps we make, the smaller the additional improvement becomes), and because it requires knowledge of a parameter of the difference of tolls to determine the suboptimal triple-step tolls (the more steps we make the more the given parameters need), we may not wish to consider more than two steps in a multi-step toll system.

4. CONCLUSIONS

We have developed the flexible single- and multi-step toll systems in which the maximum and alternative (less than the maximum) percentages of the queueing time can be saved for commuters in exchange for payment of the optimal and suboptimal tolls, respectively. The optimal and suboptimal tolls are designed to be no larger than the optimal time-varying toll for the purpose of making commuters no worse off than they would be in the nontoll equilibrium.

We have proved that the optimal single-step toll can remove at most half of the total queueing time that exists under the nontoll equilibrium. Furthermore, we have provided a rule that the optimal n-step toll can remove at most $n/(n + 1)$ of the total queueing time that exists under the nontoll equilibrium.

On the other hand, we have shown that there are two suboptimal solutions for the single-step toll system to achieve an under-half of queueing removal. One is a larger toll with a shorter charging period, and the other is a smaller toll with a longer charging period. Because both raise the same toll revenue, the authorities face a trade-off in choosing "smaller charge" or "shorter charging interval" at practical requests. Furthermore, we have derived the suboptimal double- and triple-step tolls to achieve ($\frac{1}{2} \sim \frac{2}{3}$) and ($\frac{2}{3} \sim \frac{3}{4}$), respectively of the total queueing removals.

All of these results are important references for decision-making and allow policy makers to easily compute the toll levels and starting and ending times of the tolls so as to meet different desired amounts of queueing removal without practical problems arising. Thus, it is hoped that our study can provide government or related authorities a mechanism for demonstrating potential benefits of the step toll system to the general public.

We may find that there are many people who are against the introduction of the road congestion toll for various reasons (an invasion of the right to use roads by everyone, a tax increase, etc.). Hong Kong's experiment with electronic road pricing showed that there is no technological problem in feasibility, but the authorities failed to implement this pricing due to strong opposition from the public. Therefore, if the authorities wait for the achievement of mutual understanding among citizens, they may never relieve urban traffic jams during the commuting rush hour.

Acknowledgements—The author thanks C. H. Chu of Iowa State University, J. Sicilian of the University of Kansas and the referee of *Transportation Research* for helpful comments and suggestions. The central idea for Section 3.2 of this paper was provided to the author by J. Sicilian.

REFERENCES

Arnott R. J., De Palma A. and Lindsey R. (1987) *Economics of a Bottleneck: Elastic Demand*. Discussion Paper No. 690. Queen's University, Kingston, Ontario, Canada.
Arnott R. J., De Palma A. and Lindsey R. (1990a) Economics of bottleneck. *Journal of Urban Economics*, **27**, 111–130.
Arnott R. J., De Palma A. and Lindsey R. (1990b) Departure time and route choice for the morning commute. *Transpn. Res.*, **24B**, 209–228.

Ben-Akiva M., Cyna M. and De Palma A. (1984) Dynamic model of peak period congestion. *Transpn. Res.*, **18B**, 339–355.

Ben-Akiva M., De Palma A. and Kanaroglou P. (1986) Dynamic model of peak period traffic congestion with elastic arrival rates. *Transpn. Sci.*, **20**, 164–181.

Borins S. F. (1988) Electronic road pricing: An idea whose time may never come. *Transpn. Res.*, **22A**, 37–44.

Braid M. (1989) Uniform versus peak-load pricing of a bottleneck with elastic demand. *Journal of Urban Economics*, **26**, 320-327.

Cohen Y. (1987) Commuter welfare under peak-load congestion tolls: Who gains and who loses? *International Journal of Transport Economics*, **14**, 239-266.

Daganzo C. F. (1985) The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transpn. Sci.*, **19**, 29–37.

Dawson A. L. and Brown F. N. (1985) Electronic road pricing in Hong Kong: A fair way to go? *Traffic Engineering and Control*, **26**, 522-529.

De Palma A. and Arnott R. (1986) Usage-dependent peak-load pricing. *Economic Letter*, **20**, 101-105.

De Palma A., Ben-Akiva M., Lefevre C. and Litinas N. (1983) Stochastic equilibrium model of peak period traffic congestion. *Transpn. Sci.*, **17**, 430-453.

Hau T. D. (1990) Electronic road pricing. Developments in Hong Kong 1983-1989. *Journal of Transport Economics and Policy*, **24**, 203–214.

Hendrickson C. and Kocur G. (1981) Schedule delay and departure time decisions in a deterministic model. *Transpn. Sci.*, **15**, 62–77.

Laih C. H. (1990) Traffic Problems and Policies in Taipei, Taiwan. *Presented at the Workshop of Regional Science and Urban Economics in Tsukuba University*, Tsukuba, Ibaraki, Japan.

Small K. (1982) The scheduling of consumer activities. *American Economic Review*, **72**, 467–479.

Smith M. J. (1984) The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transpn. Sci.*, **18**, 385-394.

Vickrey W. S. (1969) Congestion theory and transport investment. *American Economic Review*, **59**, 251-261.