PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Phil. Trans. R. Soc. A (2008) **366**, 1921–1930
doi:10.1098/rsta.2008.0015
Published online 6 March 2008

Congestion, ramp metering and tolls

By Pravin Varaiya*

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA

The cell transmission model of a freeway is used to compare four congestion-reducing schemes: (R) ramp control only; (T) one lane is tolled and ramps are uncontrolled; (B) bottlenecks are tolled and ramps are uncontrolled; and (RB) ramps are controlled and bottlenecks are tolled. In the base case, ramps are uncontrolled and there are no tolls. It is found that (T) is inefficient and may leave all travellers worse off; (R), (B) and (RB) can achieve efficient freeway use; (B) can eliminate queues, but has adverse spatial and equity side effects; (RB) minimizes these side effects. (RB) is likely to be least costly to implement and maintain.

Keywords: freeway congestion; ramp metering; tolls; bottlenecks; HOT lanes

1. Introduction

In November 2006, the California Department of Transportation (Caltrans) launched a programme to reduce congestion, focusing on ramp metering, incident management, traveller information and demand management (including using tolls). This paper explores the contribution that ramp metering and tolls can make, within the framework of the cell transmission model (CTM) of §2. Section 3 shows that appropriate ramp metering eliminates queue spillover, resulting in higher flows and lower travel time. Section 4 compares four strategies: (R) ramp control only; (T) one lane is tolled and ramps are uncontrolled; (B) bottlenecks are tolled and ramps are uncontrolled; and (RB) ramps are controlled and bottlenecks are tolled. In the base case, ramps are uncontrolled and there are no tolls.

2. Cell transmission model

This section is based on Gomes *et al.* (in press). The freeway is divided into N sections or cells, each with one on- and one off-ramp (figure 1). Vehicles move from right to left. Section i is upstream of section $(i-1)$. The two boundary conditions are: free flow prevails downstream of section 0, and vehicles enter the freeway at an 'on-ramp' with specified inflow r_N. The flow accepted by section $(N-1)$ is $f_N(k)$ vehicles in period k; the difference forms a queue of size $n_N(k)$. Table 1 lists the model variables and plausible parameter values, e.g. a capacity of 20 vehicles per period per lane or 2000 vehicles per hour per lane, and free flow

*varaiya@eecs.berkeley.edu

One contribution of 16 to a Discussion Meeting Issue 'Networks: modelling and control'.

 This journal is © 2008 The Royal Society

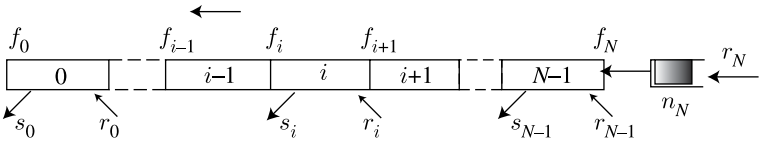

Figure 1. The freeway has N sections, each with one on-ramp and one off-ramp.

Table 1. Model variables and parameters.

symbol	name	value	unit
	section length	1	miles
	period	0.01	hours
F_i	capacity per lane	20	veh per period
v_i	free flow speed	0.6	section per period
w_i	congestion wave speed	0.2	section per period
n_i^{c}	critical density	33	veh per section
$\bar{n}_i$	jam density	133	veh per section
β_i	split ratio	$\in[0,1)$	dimensionless
$\bar{\beta}_i$	complementary split ratio $=1-\beta_i$	$\in(0,1]$	dimensionless
k	period number	integer	dimensionless
$f_i(k)$	flow from section i to $i-1$ in period k	variable	veh per period
$s_i(k), r_i(k)$	off-ramp, on-ramp flow in section i in period k	variable	veh per period
$n_i(k)$	number of vehicles in section i in period k	variable	veh per section

speed of 0.6 sections per period or 60 mph. Section length is normalized to 1 by absorbing differences in length in the speeds v_i, w_i, with $0<v_i$, $w_i<1$. The off-ramp flow is a fixed portion β_i of the total flow

$$s_i(k) = \beta_i(s_i(k) + f_i(k)) \quad \text{or} \quad s_i(k) = [\beta_i/(1-\beta_i)]f_i(k).$$

With $\bar{\beta}_i = 1-\beta_i$, the model is, for $k \geq 0$,

$$n_i(k+1) = n_i(k) - f_i(k)/\bar{\beta}_i + f_{i+1}(k) + r_i(k), \quad 0 \leq i \leq N-1, \tag{2.1}$$

$$f_i(k) = \min\{\bar{\beta}_i\, v_i\, n_i(k), w_{i-1}[\bar{n}_{i-1} - n_{i-1}(k)], F_i\}, \quad 1 \leq i \leq N, \tag{2.2}$$

$$f_0(k) = \min\{\bar{\beta}_0\, v_0\, n_0(k), F_0\}, \tag{2.3}$$

$$n_N(k+1) = n_N(k) - f_N(k) + r_N(k). \tag{2.4}$$

Flow conservation in section $i \leq N-1$ requires

$$n_i(k+1) = n_i(k) - f_i(k) + f_{i+1}(k) + r_i(k) - s_i(k), \tag{2.5}$$

which is equivalent to (2.1), using $s_i(k) = \beta_i/\bar{\beta}_i f_i(k)$; and in section N by (2.4). The flow $f_i(k)$ from section i to $i-1$ is governed by the 'fundamental diagram' (2.2). Equation (2.3) indicates that there is no congestion downstream of section 0. It is assumed that the flows $s_i(k)$ are not constrained by off-ramp capacity. The system *state* is the N-dimensional vector $\boldsymbol{n}(k) = (n_0(k), \ldots, n_{N-1}(k))$.

For fixed, stationary demands $r_i(k) \equiv r_i$, the *equilibrium flow* vector $\boldsymbol{f}(r) = (f_0, \ldots, f_N)$ is given by

$$f_N = r_N, \qquad f_i = \bar{\beta}_i(f_{i+1} + r_i), \quad 0 \le i \le N-1. \tag{2.6}$$

Demand r is *feasible* if $0 \le f_i \le F_i$, $0 \le i \le N$, *strictly feasible* if $0 \le f_i < F_i$, $0 \le i \le N$ and *infeasible* if $f_i > F_i$ for some i.

The n is an *equilibrium state* if the trajectory $n(k) \equiv n$ is a solution of (2.1)–(2.3):

$$f_i = \min\{\bar{\beta}_i\, v_i\, n_i, F_i - w_{i-1}[n_{i-1} - n_{i-1}^{\mathrm{c}}], F_i\}, \quad 1 \le i \le N-1, \tag{2.7}$$

$$f_0 = \min\{\bar{\beta}_0\, v_0\, n_0, F_0\}. \tag{2.8}$$

At equilibrium n, section i is *uncongested* if $0 \le n_i \le n_i^{\mathrm{c}}$ and *congested* if $n_i > n_i^{\mathrm{c}}$; n is *uncongested* if all sections are uncongested, otherwise it is congested. (In an uncongested section, free flow speed prevails; in a congested section, speed is lower.)

A feasible demand r has a unique uncongested equilibrium $n^{\mathrm{u}}(r)$,

$$n_i^{\mathrm{u}}(r) = (\bar{\beta}_i\, v_i)^{-1} f_i(r), \quad 0 \le i \le N-1, \tag{2.9}$$

and a continuum of congested equilibria $E = E(r)$, the solutions of (2.7) and (2.8).

Section i is a *bottleneck* if $f_i = F_i$. If there are bottlenecks at $0 \le I_1 < I_2 < \cdots < I_K \le N-1$, partition the freeway into segments $S^0, \ldots, S^K$,

$$S^0 = \{0, \ldots, I_1 - 1\}, \quad S^1 = \{I_1, \ldots, I_2 - 1\}, \ldots, \quad S^K = \{I_K, \ldots, N-1\}, \tag{2.10}$$

and the state vector $\boldsymbol{n} = (n_0, \ldots, n_{N-1})$ into sub-vectors $\boldsymbol{n} = (n^0, \ldots, n^K)$. Equations (2.7) and (2.8) partition into $1+K$ *decoupled* conditions, one for each segment, which decompose the equilibrium set into a product,

$$E(r) = E^0(r) \times \cdots \times E^K(r).$$

Since S^0 is uncongested, $E^0(r)$ consists of the unique uncongested equilibrium $n^{\mathrm{u},0}(r)$, $n_i^{\mathrm{u},0}(r) = (\bar{\beta}_i\ v_i)^{-1} f_i$, $i \in S^0$. For $k \ge 1$, the $E^k(r)$ have a similar structure, differing only in the number of sections in S^k. To explore this structure, consider a generic segment $S = \{0, \ldots, N-1\}$ with N cells, demand r and flow f with $f_0 = F_0$ and $f_i < F_i$, $i > 0$. Proposition 2.1 characterizes the set of equilibria E.

Define the *congested* equilibrium density

$$n_i^{\mathrm{con}}(r) = n_i^{\mathrm{c}} + w_i^{-1}(F_{i+1} - f_{i+1}). \tag{2.11}$$

Note that $n_i^{\mathrm{u}}(r)$ depends on the flow f_i in section i whereas $n_i^{\mathrm{con}}(r)$ depends on the immediately upstream flow f_{i+1}. Also, $n_i^{\mathrm{u}}(r) \le n_i^{\mathrm{c}} \le n_i^{\mathrm{con}}$, $i = 0, \ldots, N-1$. Denote $\tilde{n}^{-1} = n^{\mathrm{u}}$, the uncongested equilibrium (2.9), and

$$\tilde{n}^k = (n_0^{\mathrm{con}}, \ldots, n_{k-1}^{\mathrm{con}}, n_k^{\mathrm{u}}, \ldots, n_{N-1}^{\mathrm{u}}), \quad k = 0, \ldots, N-1. \tag{2.12}$$

In state $\tilde{n}^k$ the first k sections are congested, the rest are uncongested.

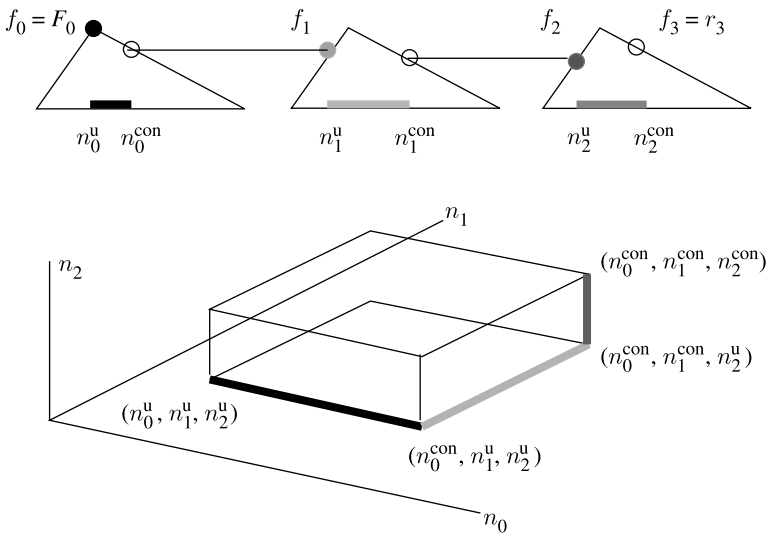

Figure 2. Fundamental diagrams and equilibrium set for example 2.2.

Proposition 2.1. *$E = [\tilde{n}^{-1}, \tilde{n}^0] \cup [\tilde{n}^0, \tilde{n}^1] \cup \cdots \cup [\tilde{n}^{N-2}, \tilde{n}^{N-1}]$, in which $[\tilde{n}^{k-1}, \tilde{n}^k]$ is the straight line segment joining $\tilde{n}^{k-1}$ and $\tilde{n}^k$. Hence, $n^{\mathrm{u}}(r) \le n \le n^{\mathrm{con}}(r) = \tilde{n}^{N-1}$, $n \in E$, with $n^{\mathrm{u}}(r)$ the unique uncongested equilibrium and $n^{\mathrm{con}}(r)$ the unique* most congested *equilibrium.*

Example 2.2. Figure 2 displays the equilibrium set E for a freeway with three identical sections and flows $f_0 = F_0$, $f_1 < F_1$, $f_2 < F_2$ and $f_3 = r_3$. Section 0 is the only bottleneck. (Assume zero off-ramp flows.) E is the union of three straight line segments. Note that congestion starts at a bottleneck and propagates upstream, as in California freeways. Suppose the freeway has the parameters of table 1, and $F_0 = 2000$ vph, $f_1 = f_2 = f_3 = 1600$ vph. Then the congested density is $n_i^{\mathrm{con}} = 53$ vehicles per lane mile with speed $1600/53 = 30$ mph. Thus, even with the *same* flow, the freeway speed at n^{u} will be 60 mph, but the speed at n^{con} will be 30 mph. The uncongested density is $n_i^{\mathrm{u}} = 1600/60 \approx 27$ vehicles per lane mile. So by reducing speed from 60 to 30 mph, the freeway 'stores' an additional $53 - 27 = 26$ vehicles per lane mile. Since one mile of a ramp can store nearly 133 vehicles (the jam density), the freeway is an expensive substitute for a ramp for storing vehicles.

3. Infeasible demand and efficient ramp metering

Peak hour demand may be infeasible and the preceding analysis needs modification. Consider example 3.1 from Gomes *et al.* (in press).

Example 3.1. Figure 3*a* shows a three-lane freeway with four identical three-lane sections. With the parameters of table 1, each section has a capacity of 6000 vph. Demand $r = (r_0 = 1200, r_1 = 0, r_2 = 2700, r_3 = 2000, r_4 = 4000)$. $\beta_0 = 0$; for $i \neq 0$, $\beta_i = \beta = 0.2$, so $\bar{\beta} = 0.8$. Denote $\alpha = \beta[\bar{\beta}]^{-1} = 0.25$. Then, r is feasible and the equilibrium flow $\phi = (\phi_0 = 6000, \phi_1 = 4800, \phi_2 = 6000, \phi_3 = 4800, \phi_4 = 4000)$. The off-ramp flow in section i is $\alpha\phi_i$. Sections 0 and 2 are bottlenecks.

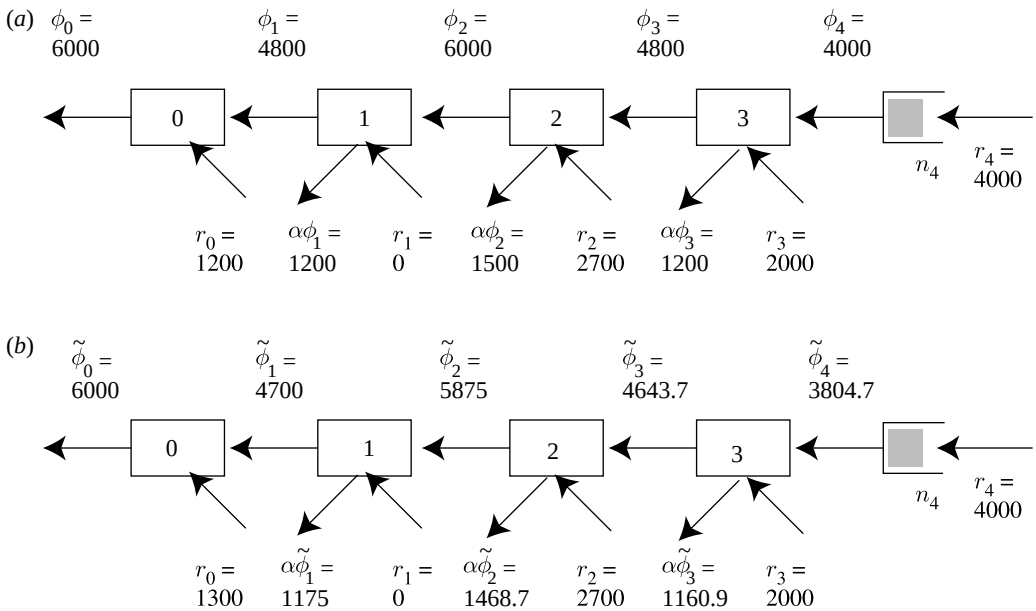

Figure 3. Example 3.1: (*a*) feasible demand and (*b*) excess demand.

Suppose demand increases to $\tilde{r}$, with $\tilde{r}_0 = 1300 > r_1$ and $\tilde{r}_i = r_i$, $i \geq 0$. The $\tilde{r}$ is *infeasible* because it would increase ϕ_0 to $\phi_1 + \tilde{r}_0 = 6100$, which exceeds its capacity. This forces a reduction in the flow out of section 1 from $\phi_1 = 4800$ to $\tilde{\phi}_1 = 4700$, from $\phi_2 = 6000$ to $\tilde{\phi}_2 = 5875$, $\phi_3 = 4800$ to $\tilde{\phi}_3 = 4643.75$ and $\phi_4 = 4000$ to $\tilde{\phi}_4 = 3804.6875$. So $n_4(k)$ will grow at the rate of $4000 - 3804.6875 = 195.3125$ vph. All sections will become congested, with the new equilibrium flows of figure 3*b*.

For the model of §2, let ϕ be the solution of (2.6): $\phi_N = r_N$, $\phi_i = \bar{\beta}_i(\phi_{i+1} + r_i)$, $0 \leq i \leq N-1$. Suppose r is *infeasible*, so that $\phi_i > F_i$ for some i. To simplify the notation assume that $\phi_0 > F_0$. Also assume that the demand becomes feasible if either $r_N = 0$ or $r_0 = 0$. Let

$$\tilde{r}_N = \max\{\rho \geq 0 | \text{ the demand } (r_0, \ldots, r_{N-1}, \rho) \text{ is feasible}\}, \tag{3.1}$$

$$\hat{r}_0 = \max\{\rho \geq 0 | \text{ the demand } (\rho, r_1, \ldots, r_N) \text{ is feasible}\}. \tag{3.2}$$

Proposition 3.2. (*i*) *$\tilde{r}_N < r_N$ is the largest upstream flow for which the demand $\tilde{r} = (r_0, \ldots, r_{N-1}, \tilde{r}_N)$ is feasible. The corresponding equilibrium flow $\tilde{\phi}$ is*

$$\tilde{\phi}_N = \tilde{r}_N, \qquad \phi_i = \bar{\beta}_i(\tilde{\phi}_{i+1} + r_i), \quad 0 \leq i \leq N-1.$$

(*ii*) *With demand r, under the no-metering strategy the system converges to the most congested equilibrium $n^{\text{con}} \in E(\tilde{r})$. The queue $n_N(k)$ at the upstream ramp grows at rate $(r_N - \tilde{r}_N)$.* (*iii*) *$\hat{r}_0 < r_0$ is the largest flow for which $\hat{r} = (\hat{r}_0, r_1, \ldots, r_N)$ is feasible. The equilibrium flow $\hat{\phi}$ is*

$$\hat{\phi}_N = r_N, \qquad \hat{\phi}_i = \bar{\beta}_i(\hat{\phi}_{i+1} + r_i), \quad 1 \leq i \leq N-1, \quad \hat{\phi}_0 = \bar{\beta}_0(\hat{\phi}_1 + \hat{r}_0).$$

Under the ramp-metering strategy that reduces the on-ramp flow from r_0 to $\hat{r}_0$, the system converges to an equilibrium in $E(\hat{r})$. The queue at the on-ramp in section 0 grows at rate $(r_0 - \hat{r}_0)$. (iv) The ramp-metering strategy yields larger flows:

$$\tilde{\phi}_i < \hat{\phi}_i, \quad 1 \le i \le N \quad \text{and} \quad \tilde{\phi}_0 = \hat{\phi}_0 = F_0.$$

If $\beta_i > 0$ for some $i \ge 1$, the total discharge under the ramp-metering strategy is strictly larger than under the no-metering strategy, and

$$\mu = \frac{r_N - \tilde{r}_N}{r_0 - \hat{r}_0} = (\bar{\beta}_1 \ldots \bar{\beta}_N)^{-1} > 1. \tag{3.3}$$

Substituting the split ratios of example 3.1 into (3.3) gives the ramp-metering 'gain' $\mu = (\bar{\beta}_1 \bar{\beta}_2 \bar{\beta}_3)^{-1} = (0.8)^{-3} \approx 2$. The next result is not difficult to prove.

Proposition 3.3. *There is a ramp-metering strategy that achieves the metering gain. By keeping flow strictly below the capacity, the strategy achieves the uncongested equilibrium.*

4. Ramp metering, lane tolls and bottleneck tolls

We focus on example 3.1 with fixed, infeasible demand. The no-metering, no-tolls *base case* leads to the most congested equilibrium with user equilibrium flows of figure 3*b*; speed is 30 mph and the queue at the freeway entrance grows at a rate of 196 vph. We consider four congestion reduction schemes.

(*a*) (*R*) *Ramp metering*

By proposition 3.3, ramp metering can achieve efficient use of the freeway, reducing unfulfilled demand to 100 vph. At a small capacity loss, the system converges to the uncongested equilibrium with a speed of 60 mph. There is a deadweight welfare loss from queuing delays at the metered ramps.

(*b*) (*T*) *Tolling one lane*

In this scheme, one lane is tolled, the two 'free' lanes are not metered. The scheme is similar to those used in I-15 and SR-91 in California.

First consider the simple case of a freeway with a single inflow of $\phi_4 = r$ vph and a single outflow of $\phi_0 = r$ vph (figure 3). Suppose, as in a traditional model, that travel time is an increasing convex function $\tau(\rho)$ of the flow/capacity ratio ρ. Following Small & Yan (2001), it is easily seen that all travellers are worse off than in the base case, except for those with a very high value of travel time.

Now consider the CTM of example 3.1. Suppose one lane is tolled in such a way that tolled travellers experience no congestion, travelling at 60 mph. Demand on the free lanes will be infeasible, traffic on the free lanes will settle to the most congested equilibrium, and these lanes will become congested throughout their length. Thus *the tolled lane must extend all along the freeway.* This is a waste of the freeway capacity, because as is clear from figure 3 the capacity in all sections of the freeway except section 0 is underused. Again, all travellers (except for those with a very high value of travel time) will be

worse off. Moreover, the demand on the free lanes will be greater than in the base case, and so the unfulfilled demand will be *larger* than the unmet demand of 196 vph in the base case.

(*c*) (*B*) *Bottleneck toll*

Since in example 3.1 the only bottleneck is section 0, results about pricing to achieve efficient bandwidth allocation in communication networks (Kelly 1997) suggest that only vehicles entering section 0 should be tolled. Kelly's proof does not apply directly because the communication network model is different from the CTM. However, a proof that only bottlenecks need to be tolled can be based on the following observations. (i) If section 0 alone is tolled, an appropriate toll will reduce demand destined for section 0 from 6100 to 6000 vph. This reduction may not entirely be borne by the flow $\tilde{r}_0$ as was achieved by ramp metering. The freeway will be efficiently used. (ii) A toll in any other section will reduce demand exiting before section 0, which is inefficient. (iii) If demand is elastic, there will be an increase in vehicles exiting other sections, resulting in a welfare 'bonus'.

The demand for trips through section 0 is $r = 6100$ vph and capacity is 6000 vph. So the toll should be so large as to decrease the demand by the factor

$$\frac{\Delta r}{r} = \frac{6100 - 6000}{6100} \approx \frac{1}{60}.$$

If P is the cost of a trip through section 0, $r(P)$ is the aggregate demand function, and $-\epsilon < 0$ is its elasticity, the toll ΔP that reduces the demand by Δr is

$$\Delta P = \frac{1}{\epsilon}\frac{\Delta r}{r} \times P. \tag{4.1}$$

If this toll is imposed, demand will decrease to 6000 vph. If demand is reduced slightly further, the system will converge to an uncongested equilibrium and vehicles will move at free flow speeds throughout. Moreover, the reduction in demand will eliminate ramp queues. The toll (4.1) not only achieves efficient use of the freeway but also eliminates the deadweight loss of queuing delay. Thus, the bottleneck toll seems ideal. However, implementation difficulties and adverse spatial and equity side effects reduce the attractiveness of bottleneck tolls.

The typical California freeway has several bottlenecks and imposing a toll on each of them will make implementation difficult.

Motorists will avoid bottleneck tolls by exiting the freeway before the tolled sections. This could 'move' the bottleneck upstream or create an additional bottleneck. Furthermore, the streets leading from these exits will carry more traffic, which may create new congestion and public opposition.

To understand the equity side effect, we evaluate what the toll is likely to be. Suppose $P = 20$, and $\epsilon = 1/3$, within the range suggested in Small & Yan (2001). To reduce the demand from 6100 to 6000 vph the toll (4.1) should be $3 \times 1/60 \times 20 = \$1$. *Every one* of the 6000 drivers will pay this toll so that the total toll revenue will be \$6000 per hour, which amounts to \$60 for each of 100 diverted vehicles. Small & Yan (2001) estimate $\epsilon = 1/3$ for the case when motorists unwilling to take the tolled lane could use an adjacent free lane, which is a close substitute. When everyone travelling through a bottleneck is tolled, the alternatives open to someone unwilling to pay the toll are worse, so the elasticity

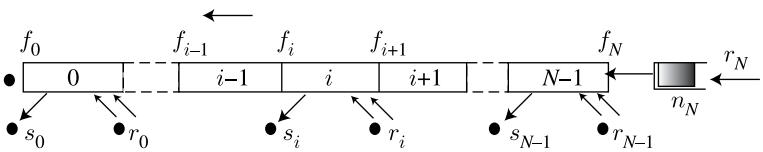

Figure 4. Each section has two on-ramps, one tolled and one metered, and one off-ramp (circles, tag reader).

will be lower, and the toll correspondingly higher. Furthermore, if the 'excess' demand increases to 200 vph, the toll (4.1) will increase to \$2. Many of the 6000 toll-paying motorists will be worse off than in the base case, making public acceptance of the bottleneck toll unlikely. Of course the tolling authority will collect large revenues.

(*d*) (*RB*) *Ramp metering and bottleneck toll*

This scheme can eliminate the adverse spatial and equity side effects of the pure bottleneck toll scheme. The scheme imposes ramp metering to ensure that the freeways operate efficiently and vehicles move at free flow speed. This will create queues at ramps. The queue can be bypassed by paying a toll.

The scheme is illustrated in figure 4, which is similar to figure 1, except that there are two on-ramps in each section, of which one is metered and the other is tolled. The tolled on-ramp is equipped with an electronic tag reader. Every off-ramp is also equipped with a tag reader. Vehicles entering a tolled ramp must carry a tag. The tag reader registers the entry time, vehicle tag ID and the entry ramp ID. When the vehicle leaves the freeway from an off-ramp, its tag reader registers the vehicle ID and the exit ramp ID. The vehicle is charged a toll that depends on the time, entry ramp, exit ramp and, possibly, traffic conditions.

The ramp metering ensures that the freeway operates efficiently. Motorists who enter a metered ramp will wait in the ramp queue, but pay no toll. By paying a toll a motorist can bypass the queue at the ramp.

The toll depends on the entrance and exit ramps. The toll should reflect the bottleneck sections crossed between the entrance and exit ramps. This affords great flexibility. Consider example 3.1, in which section 0 is the only bottleneck. In this case, ideally, only those tolled vehicles that entered through a tolled ramp and exit from section 0 will be charged. A tagged vehicle entering a tolled ramp but leaving before section 0 pays no toll. A tagged or untagged vehicle that enters through a metered ramp also pays no toll. If this encourages drivers to exit earlier, creating the spatial side effect noted above, a smaller toll could be charged for leaving, say, in section 1. This flexibility can be used to reduce or eliminate the spatial side effect of scheme (B). On the other hand, drivers wishing to go through section 0 without paying a toll can use the metered ramp. They will incur a slightly larger delay than under scheme (R) owing to toll-paying vehicles going through section 0, but real and perceived inequity will be lower.

When the freeway has several bottlenecks—identified by flows close to capacity—there is a toll for each bottleneck and, upon exit, a tolled vehicle is charged the sum of the tolls at all bottlenecks it has traversed. The tolls should be larger than those under scheme (B). The larger the tolls, the more the scheme favours those using metered ramps and those exiting before any bottleneck.

The distributional impact of (RB) is worth analysis. Also worth investigation are practical schemes for setting tolls. Trying to do so through estimating demand elasticities as in (4.1) seems fruitless considering the time-varying nature of the demand. A better approach may be an adaptive scheme in which the toll is increased until flow at bottlenecks is below capacity, together with some agreement about favouring toll-paying versus time-paying travellers.

In the (T) scheme, the tolling authority has an incentive to increase congestion on the free lanes. By contrast, in the (RB) scheme, the tolling authority has the incentive to meter the ramps more aggressively than necessary in order to create a larger ramp queue, encouraging more motorists to pay a toll. However, this will lead to underuse of the freeway that is immediately perceived by motorists.

One disadvantage of the (RB) scheme is the requirement of two on-ramps, for which there may be no space. However, it is not necessary to have both types of ramps in each section. Some entrances may have a single-tolled ramp; others may have a single-metered ramp.

The costs of the electronic tag readers on ramps should be much less than implementing a tolled lane (scheme (T)), requiring a dense set of tag readers along a freeway, together with some physical barrier separating the tolled lane from the free lanes. Equally important is the fact that the (RB) scheme reverts to (R) simply by setting tolls to zero, and to the base case by not metering the ramps, whereas with its physical barrier, the (T) scheme imposes a costly-to-reverse change in the road infrastructure. The cost of the (RB) scheme will also be less than (B), which requires tag readers across all lanes of all bottleneck sections. The cost of enforcement of the (RB) scheme is also much lower than scheme (T) or (B).

5. Conclusions

The CTM is used to analyse ramp metering and congestion tolls—two congestion reduction measures favoured by California. Four schemes are considered: (R) relies exclusively on ramp metering; (T) tolls one lane in a multilane freeway, with no metering on the free lanes; (B) tolls bottleneck sections; and (RB) combines ramp metering with bottleneck tolls. In the base case, there is no metering and no tolls.

Metering algorithms can be designed so that (R) achieves efficient use of the freeways, keeps vehicles moving at free flow speeds, and achieves lower total travel time than in the base case. However, delay at the ramps constitutes a deadweight welfare loss.

The I-15 freeway in San Diego and SR-91 in Orange County, which combine tolled lanes with adjacent free lanes, are schemes of type (T). Under (T), travellers on both the tolled lanes and on the free lanes are worse off than in the base case. (Travellers with a much higher value of time than the average might be better off.) The freeway is inefficiently used, and the tolling authority has the incentive to worsen congestion on the free lanes. Thus, from a public policy viewpoint, one cannot justify scheme (T).

Pricing bottlenecks (B) can simultaneously keep freeways efficient and eliminate the deadweight queuing delay at ramps. However, this scheme suffers from adverse spatial and equity side effects that will make public acceptance unlikely.

The (RB) scheme, combining ramp metering and bottleneck pricing, avoids these adverse side effects. It offers a very flexible means to manage the freeway demand. Its cost of implementation is much lower than scheme (T) or (B), and it creates the incentive for the tolling authority to operate the freeway efficiently. It combines the best aspects of both (R) and (B) schemes.

This is a report of the Tools for Operations Planning (TOPL) Development Group, whose work is supported by the California Department of Transportation through the California PATH Program. The contents of this paper reflect the views of the author who is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policy of the California Department of Transportation. This paper does not constitute a standard, specification or regulation. I am grateful to Hani Mahmassani for suggesting the (RB) scheme, to Frank Kelly for critical comments, and to Vassili Alexiadis of Cambridge Systematics and members of the TOPL group for their discussions. A version of this paper was presented at the Twelfth International Conference of Hong Kong Society for Transportation Studies, December 2007.

References

Gomes, G., Horowitz, R., Kurzhanskiy, A. A., Varaiya, P. & Kwon, J. In press. Behavior of the cell transmission model and effectiveness of ramp metering. *Transport. Res. Part C.* (doi:10.1016/j.trc.2007.10.005)

Kelly, F. P. 1997 Charging and rate control for elastic traffic. *Eur. Trans. Telcommun.* **8**, 33–37.

Small, K. A. & Yan, J. 2001 The value of 'value pricing' of roads: second-best pricing and product differentiation. *J. Urban Econ.* **49**, 310–336. (doi:10.1006/juec.2000.2195)