On the Impact of HOT Lane Tolling Strategies on Total Traffic Level

by Soheil Sibdari and Mansoureh Jeihani

This paper shows how tolling (or pricing) strategies can be used to control the congestion levels of both untolled and high occupancy toll (HOT) lanes. Using a user-equilibrium method, the paper calculates the number of travelers on each route during the peak period and provides a numerical analysis that determines the distribution of travelers for different tolling strategies. It shows that with the right tolling strategy some travelers who initially plan to use the untolled lane during the peak period will change both their routes (i.e., select the HOT lane) and departure times (i.e., depart earlier or later). Using this result, the paper compares static and dynamic pricing strategies and shows that with a dynamic strategy a larger profit can be earned and congestion reduced in the untolled lane.

INTRODUCTION

Traffic congestion is a serious problem in urban areas, challenging transportation policymakers and raising the total cost of transportation. Traffic demand management has become an increasingly popular tool for managing congestion. This approach assumes travel demand is a quantity requested by road users whose size and distribution can be controlled by different policies. These policies include congestion pricing, flexible work hours, carpooling, and public transportation. Many congestion pricing studies have developed policies that simply maximize expected profit or maintain an acceptable level of congestion for a high occupancy toll (HOT) lane. However, many transportation planners are more concerned with reducing congestion levels than maximizing profits, and would benefit from a pricing policy that changes a traveler's behavior and significantly lowers overall congestion.

Some researchers, including Small and Yan (2001) and Fielding and Klein (1993), have addressed the concept of value pricing, which lets travelers choose between a free but congested lane and a priced but free-flowing roadway. Also called HOT lanes, these priced roadways allow single drivers to pay for more highly valued services. HOT lanes currently in use in the United States include Interstate 15 in San Diego, Quick Ride System on the Katie Highway and the Northwest Freeway in Houston, and the SR 91 Express Lanes in Orange County, California (Evans et al. 1993).

Value pricing can also solve some of the problems associated with high occupancy vehicle (HOV) lanes, such as when they become congested from vehicles with two passengers (HOV-2), while those HOV lanes allowing vehicles with at least three passengers (HOV-3) are underutilized. HOT lanes allow single occupant vehicles (SOVs) to enter a lane by paying a toll, while carpools or buses can use the same lanes for free or at a discounted rate. Allowing SOVs to use HOV lanes via an appropriate toll can control congestion and generate profit for the toll-road operators.

The existing pricing literature considers profit maximization, second-best pricing, social welfare maximization, and a minimum level of service requirement as objectives in their model development. A second-best pricing policy maximizes social welfare subject to a zero-toll constraint on the alternative roadways. The optimal solution for this policy is a weighted average of the marginal external congestion costs between non-carpooling and carpooling vehicles. Profit (or revenue) maximization allows a planner to maximize profit subject to a zero-toll constraint on other roads. Small and Yan (2001) compared profit maximization and revenue maximization and found that travelers' behaviors under both policies are almost identical. They also compared the outcomes of profit maximization and second-best pricing, and showed that profit maximization sets higher

tolls and lowers total social benefits. Another type of pricing regime, maintaining service level, sets the toll high enough to keep the flow of the priced roadway at a minimum specified speed. Similar to second-best pricing, this objective achieves social optimum and has an additional constraint that guarantees a minimum level of service in the express lane. A detailed review of these pricing policies is provided by Yang and Huang (1999), Braid (1996) and Verhoef et al. (1996).

Unlike the literature above, this paper examines the effect of HOT lane pricing on the congestion level of a regular lane, i.e., untolled lane. It investigates a situation where a regular lane is congested during the peak (i.e., 5-6 p m.) and the HOT lane is under-utilized during the shoulders of the peak period (i.e., 3-4 p.m. or 7-8 p.m.). It shows that an appropriate toll motivates some travelers who plan to use the untolled lane during the peak period to use the HOT lane during the shoulders of the peak. To illustrate, consider a regular and a HOT lane of the same length and free-flow traffic times connecting two points. The toll for the HOT lane is $3 from 4-5 p.m. and $6 from 5-6 p.m. A traveler, who plans to depart at 5 p.m. and is not willing to pay $6, might change her travel time to 4 p.m. to take advantage of the HOT lane's light traffic and $3 rate. This behavior smoothes out the departure rate during the peak period and lowers the regular lane's traffic.

LITERATURE REVIEW

Most studies on value pricing use either static or dynamic models. Static pricing assumes that travel demand and costs are not time sensitive. As a result, the inter-temporal impact of tolls on long-term congestion levels is not considered. Many studies in this area use marginal cost pricing to estimate an optimal congestion price on transportation networks. This pricing policy makes peak period travel more expensive with a toll that is calculated as the difference between marginal social and private costs.

Dafermos and Sparrow (1971) used static congestion pricing to determine the optimal toll for general transportation networks. Yang and Meng (1998) calculated static congestion pricing using the marginal social and private costs of travel, where roads are modeled as bottlenecks. Unlike static congestion pricing, dynamic pricing methods affect travel demand, travel costs, and toll levels over time. A seminal paper by Vickrey (1969) introduced a dynamic congestion-pricing model for a single bottleneck with a fixed capacity and number of travelers. He took into account work-related trips and assumed the same desired departure times for all travelers. However, some travelers could change their departure times to avoid traffic congestion. If a traveler left early, she faced no congestion, but could possibly encounter costs associated with early arrival. If she left to arrive on time, she could also face congestion that would make her late. Vickrey (1969) showed that in these situations equilibrium congestion occurred when no driver could reduce her trip cost by changing her departure time.

Arnott et al. (1993) extended Vickrey's model by considering heterogeneous travelers. They developed a deterministic mathematical model to establish the efffect of an optimal time-varying toll on social welfare. They also compared the impact of different tolling schemes, such as uniform and step-functions on system efficiency. With a time-dependent congestion toll and optimal congestion pricing, they achieved a more uniform departure rate and reduced both congestion cost and traveling time for commuters.

Arnott et al. (1990) also considered fixed demand in a network with parallel routes, and used dynamic traffic assignment to examine the impact of different pricing regimes on network congestion and reproduced hypercritical flow conditions. Carey and Srinivasan (1993) used non-linear programming to address system marginal costs, user externality costs, and optimal congestion tolling and developed optimal tolls under a time-dependent travel demand. They compared optimal dynamic and static tolls and found that optimal dynamic tolls depended on the congestion level and whether the congestion was at the beginning or ending of a peak period. Yang and Huang (1999) developed a time-varying pricing model for a road bottleneck when demand is elastic and deterministic. They used a continuous-time optimal control approach to maximize social benefit.

Liu and McDonald (1999) used economic and simulation models to compare first-best, second-best, and no-toll policies in a model with two routes and periods (peak and pre-peak). They found that second-best pricing policies are effective in reallocating traffic volumes, but less effective than first-best tolling. The toll level in the second-best pricing policy is lower than in a first-best tolling policy, and the social welfare benefit obtained from the second-best tolling policy is smaller than the gains from a first-best tolling policy.

All the papers reviewed manage traffic on a route by pricing it to maximize expected revenue or social welfare and make route demand (i.e., the number of passengers that use the route) solely a function of the price, thus neglecting the linkage between price and demand on alternative routes. To fill this gap, this paper addresses this linkage and uses a traffic assignment method that considers the prices charged on alternative routes. The problem is modeled from the perspective of a social planner whose objective is to control congestion. A prime factor impacting our model is the presence of travelers who adjust their departure times with respect to a route's price level. Considering this factor, the planner can manage total congestion by charging approppriate tolls. The section below addresses the problem and presents the traffic assignment method used. After that, a numerical analysis is conducted to determine the distribution of travelers during the peak period under different pricing policies followed by conclusions.

MODEL

Consider two roadways that connect points A and B. They have the same length, L, and free-flow travel time, $FFTT$. One is toll-free (NT) and the other is tolled (T). The only points of access and egress for both roadways are A and B, and travelers can only choose between these two routes to travel from A to B. HOVs can use both roadways for free, but SOVs can only use NT for free and must pay to use the HOT lane.

Next, consider a total of N travelers in SOVs with home-based-work (HBW) trips who plan to travel from A to B during the peak period. HBW trips, usually made by SOVs, are most important during peak periods because they are less flexible in departure times since each person must arrive at work at a certain time. The length of the peak period is n, and it is divided into n time slots with unit lengths. Without loss of generality, consider an odd number of periods from 3-8 p.m., giving five time slots, 3-4, 4-5, 5-6, 6-7, and 7-8 respectively.

Each traveler has a preferred departure time, and since only HBW trips are considered, assume that all travelers prefer a specific time slot, say time slot zero. Depending on the congestion level and toll, travelers might change their departure times if those times increase their disutilities of travel. Next, assume $D = [D_{-2}, D_{-1}, D_0, D_1, D_2]$ is the vector of disutilities associated with each time slot, where D_i determines the disutility of traveling in time slot i, for i = -2,-1,0,1,2. The size of this disutility is the same for all travelers, but varies by departure time. For example, departing in time slot zero imposes no disutility, while traveling in the other time slots imposes positive disutility, with the disutility of traveling in later time slots being greater than traveling in earlier time slots. However, the disutility of traveling in time slot -1 is not necessarily equal to the disutility of traveling in time slot 1. This is also true for time slots -1, 2, and 2. In addition to disutility, travel time and the toll level affect a traveler's cost, and travelers choose a roadway and time slot to minimize their costs.

To analyze the problem described, the paper relies on a traffic assignment model to find an equilibrium solution that specifies the number of travelers on each roadway and in each time slot. A wide range of traffic assignment models can be employed to assign traffic flow between an origin and a destination among different routes, with system optimal and user equilibrium being the most popular. System optimal models find an assignment that minimizes total network travel time based upon the assumption that travelers cannot change their route without increasing total system travel time. In the user equilibrium model adopted in this study, travelers cannot improve their travel time by switching routes. Using the Frank and Wolfe (1956) algorithm, all flows are assigned to the initial shortest path and the links are iteratively updated using a volume delay function, which shows the

relationship between the cost of traversing a link and the flow on it. The algorithm then finds a new shortest path between each origin and destination and assigns a convex combination of flow to the new and old shortest paths.

The above traffic assignment method is used to calculate the demand for each time slot. For a given price menu, the equilibrium number of travelers on each route where no travelers have an incentive to change their departure time or their route choice is determined. By changing price, the distribution of passengers among these time slots changes because some passengers change their routes or departure times. This change allows the model to determine the impact of each roadway's price on travel on the other roadways. Several factors impact a traveler's route choice, such as travel cost, costs of using alternative routes, discomfort associated with a route, or time of travel. The cost of travel includes out-of-pocket costs (e.g., gas, parking fees, and tolls). Since the study is on the effect of tolls and time of travel on travel behavior, the effects of gas price and parking fees are not considered. For discomfort, it is assumed that time slot 0 has no disutility, and the disutilities of the other time slots increase with distance from this slot.

To apply the algorithm, the links and their generalized cost functions are defined. Each time slot (T and NT) is treated as a separate link, and the volume delay function for each time slot i and roadway j is defined as follows.

(1) $\quad C_i^j(V_i^j) = \delta p_i^j + \gamma D_i + \varphi t_i^j [1 + \alpha (\frac{V_i^j}{Cap_i^j})^\beta]$

for $i = -(n-1)/2, \cdots, 0, \cdots, (n-1)/2$, and $j = T, NT$.

where:

C_i^j : generalized cost of travel on each link.

D_i : cost associated with the disutility of choosing time slot i rather than time slot zero. Note that the values of D_i are the same for both roadways.

t_i^j : free-flow travel time (in minutes), which is calculated as follows.

$$t_i^j = \frac{L}{FFS_i^j} * 60.$$

Cap_i^j : capacity during time slot i in roadway j.

p_i : toll level during time slot i in roadway T.

V_i^j : the flow during time slot i in roadway j.

L : length of the links (in miles).

FFS_i^j : free-flow speed of the links (in miles per hour).

$\alpha, \beta, \delta, \gamma$, and φ are the parameters of the model.

If a road is so congested during time slot i that its travel time is greater than the length of that time slot, then all the vehicles cannot clear that slot. In this situation, the overflowing vehicles will enter the next time slot. The flow of each link, V_i^j, can then be calculated using:

(2) $\quad V_i^j = \overline{V}_i^j + O_{i-1}^j$

where $\overline{V}_i^j$ is the initial flow in time slot i on roadway j, and O_{i-1}^j is the overflow of traffic from time slot $i-1$ to time slot i on roadway j. It is assumed that for $O_{-3}^j = 0$ all and the following equation is used to calculate O_i^j:

(3) $O_i^j = Cap_i^j (\dfrac{TT_i^j \quad FFTT_i^j}{\alpha FFTT_i^j})^\beta$

The volume delay function in Equation 1 is the Bureau of Public Roads (BPR) function with generalized cost, which uses travel time as the cost of traversing a link. This link travel time can be calculated using the free-flow travel time of a link and the ratio of link flow to capacity. However, the BPR function with generalized cost assumes that traversing a link conveys costs other than link travel time. These costs are presumed to be the possible toll and disutility of selecting a time slot over time slot 0. The applied Frank-Wolfe algorithm can be explained using the following procedure:

1. Initialization: Perform all-or-nothing assignment based on $C_i^j(0)$. This gives the flow vector $V_i^j(1)$. Set the iteration number m to 1.

2. Update Travel Cost: Update the link travel cost $C_i^j(m) = C_i^j(V_i^j(m))$.

3. Direction finding: Perform all-or-nothing assignment with the updated $C_i^j(m)$. This gives the auxiliary flow vector $Y_i^j(m)$.

4. Line Search: Find the solution to $\min \sum_{ij} \int_0^{V_i^j(m)+\lambda(Y_i^j(m)-V_i^j(m))} C_i^j(u)du$.

5. Move: Set the flows to $V_i^j(m+1) = V_i^j(m) + \lambda(Y_i^j(m) \quad V_i^j(m))$.

6. Convergence Test: If $\lambda < \varepsilon$ stop; otherwise go to Step 2, where ε is a very small number (e.g., 0.0001).

The algorithm yields an equilibrium solution in which travelers are distributed among the $2n$ links with minimized possible cost and all links have the same cost. Since each link represents a time slot and a roadway, the N travelers are distributed among each time slot on a toll-free or tolled lane so that they experience the same travel cost.

NUMERICAL STUDY

The simulation model was validated and calibrated using the Trans-CAD platform. The calibrated model provided information on total congestion on each road in each scenario. Using this information, total travel time, total cost of travelers, and the revenue generated by the tolls were calculated. A total of $N = 20,000$ travelers are assumed to travel from an origin to a destination by the regular and HOT lanes during the five-hour period noted. As in the previous section, there are 10 lanes of the same length ($L_i^j = 10$ miles), capacity ($Cap_i^j = 2250$ vehicles per hour per lane), and free-flow speed ($FFS_i^j = 65$ miles per hour) for all time slots and roadway combinations. Each lane is assigned to a time slot and each traveler has 10 route choices. For example, the first lane can be considered as time slot 3-4 p m. in the regular lane.

Calibration and Validation of Simulation Model

To accurately perform the simulation, the model's parameters are estimated by calibration. This technique develops travel demand by estimating various parameters and examining the ability of the model to replicate actual traffic patterns.[1] This is done by solving the equations for the parameters of interest by supplying the observed values of the variables from surveys and experimental analysis and by trial and error to find the values of the parameters with the largest probability of being accurate within an acceptable margin of error (i.e., within 0.5% of actual values). After satisfactory parameters are obtained, the model's performance is checked by comparing the observed and simulated travel times and traffic counts.

The paper also uses the calibration results in Jeihani et al. (2008). These authors used a corridor of Interstate 83 in Baltimore, Maryland, and calibrated demand parameters iteratively by comparing the outputs of their models with observed data. In each iteration, the parameters were adjusted to replicate the observed condition. Then, following Oketch and Carrick (2005), they used a modified Chi-square test based upon the difference between the simulation model's output and observed traffic counts to test their results. At the 5% level of significance, they could not reject the hypothesis that their model adequately simulated the traffic flow patterns. Similarly, to estimate travel time parameters and disutility levels, the results of Jehani et al. (2008) are used and we assume that $\alpha = 0.5$, $\beta = 4$ and the disutility vector is $D = [28,20,0,15,250]$.

The demand parameters, including δ, φ, and γ are estimated following the approaches in Litman (2008) and Kumar et al. (2004), which are based on arc elasticities of demand. Also, using the same variables as in these authors' works, the estimated generalized cost function is as follows:

$$(4) \quad C_i^j(V_i^j) \quad 4p_i^j + 0.25D_i + 0.8t_i^j[1 + 0.5(\frac{V_i^j}{Cap_i^j})^4]$$

Equation 4 is used for the rest of the numerical study.

Base Cases

To determine the initial distribution of travelers among different time slots, the first consideration is what happens when the HOT lane does not exist (No-HOT-Lane). The next is to consider when a HOT lane is available for free, and all travelers, including SOVs, can use the extra lane (Free-HOT-Lane). Given that during the peak period (3-8 p.m.), more passengers prefer to depart in the middle time slots (e.g., 5-6 p.m.), it seems reasonable to expect normal distribution of travelers in these cases. This distribution enables us to study the impact of only departure time on route choice. For example, in the case of No-HOT-lane, the consumers' generalized cost function consists of departure time disutility and travel time since travelers can choose among five time slots for the regular lane. The results are presented in Table 1. The first set of columns shows road type and time slots, and the second and third sets are the results for the No-HOT-Lane and Free-HOT-Lane scenarios. Further, the table is divided into two sets of rows, one for the regular lane and the other for the HOT lane. Four parameters are reported for each time slot: number of vehicles (flow), volume-capacity ratio (VOC), travel time (TT), and average speed of vehicles (speed).

As expected, the distribution of travelers among time slots follows normal distribution, with more travelers departing in time slot 0 (5-6 p m.). In the No-HOT-lane case, all time slots in the regular lane are over-utilized, with the volume-capacity ratio ranging from 1.68 to 2.3. The length of each time slot of 60 minutes is exceeded in all the time slots, which means that some cars overflow into the next time slot. For example, in time slot -2, where the flow is 3,772 vehicles per hour and travel time is 91.4 minutes, the road will not be cleared for travelers who depart from 3-4 p m., and some of them will continue their trips into time slot -1. Using Equation 2, the total overflow of time slot -2 is 495 vehicles, and this has been added to $\overline{V}_{-1}^{NT}$. Thus, $V_{-1}^{NT} = \overline{V}_{-1}^{NT} + O_{-2}^{NT} = 3,935+ 495 = 4,430$. The over-utilized and overflow time slots are highlighted with light and dark shadows, respectively.

Based on the length of each time slot and the estimated parameters, the free flow travel time is 18.46 minutes. In Table 1, the travel times of all time slots exceed 60 minutes, with the highest being time slot zero (155 minutes). The average speed of vehicles is also as low as 8 miles per hour while the free flow speed is 65 miles per hour. Note that the disutilities of using different time slots are $D = [28,20,0,15,25]$, which means departing earlier induces higher disutilities than departing later. The values of the disutilities give a distribution skewed toward the left (higher flows in slots 1 and 2 compared with slots -1 and -2 even without considering the overflows).

Table 1: Flow, Travel Time, and Speed of Time Slot

Road		No-HOT-Lane				Free-HOT-Lane			
Speed	Period[a]	Flow	VOC[b]	TT[c]	Speed	Flow	VOC	TT	Speed
Regular Lane	-2 (3-4)	3772	1.68	91.40	13.13	915	0.40	18.79	64.46
	-1 (4-5)	4430	1.97	104.84	11.45	2056	0.91	24.90	48.19
	0 (5-6)	5179	2.3	154.74	7.75	3001	1.33	47.66	25.18
	1 (6-7)	5069	2.25	114.94	10.44	2351	1.05	29.47	40.72
	2 (7-8)	4605	2.05	96.48	12.44	1683	0.76	21.49	55.84
HOT Lane	-2 (3-4)	-	-	-	-	920	0.41	18.68	64.34
	-1 (4-5)	-	-	-	-	2050	0.91	24.82	48.34
	0 (5-6)	-	-	-	-	3004	1.34	47.80	25.10
	1 (6-7)	-	-	-	-	2354	1.05	29.51	40.66
	2 (7-8)	-	-	-	-	1673	0.75	21.51	55.84

[a] There are five time slots for each road. The first time slot is 3-4 p m., and we label it as time slot -2. The other time slots are labebed accordingly.

[b] If $VOC_i \geq 0.9$ (or the flow of a time slot exceeds 1835), the time slot i is considered over-utilized (indicated by light shadow).

[c] If $TT_i \geq 60$, then the road i has an overflow, (indicated with dark shading).

For the rest of our numerical analysis, HOT lane is added and the results compared in terms of the impact of different pricing scenarios on total traffic level. First, consider a benchmark scenario where the HOT lane is available for free, i.e., $p = 0$. When SOVs can use the HOT lane for free, it can be treated as a newly constructed regular roadway. The second set of columns in Table 1 shows the traffic flow, VOC, travel time, and speed among five time slots of both lanes. A roadway is over-utilized if $VOC \geq 0.9$. Since the HOT lane is free, roadway use is almost even, with three over-utilized time slots (1, 0 and -1) and two under-utilized time slots (2 and -2). The traffic flow is still normally distributed, and more travelers use time slot zero than the others. Since we assumed that the disutility of departing earlier is higher than later, the distribution is skewed toward the left, which means that $V_1^j > V_{-1}^j$ and $V_2^j > V_{-2}^j$ for $j = T, NT$. Because of low demand for time slots 2 and -2, the travelers can enjoy the free-flow travel time of 18 minutes, while the travel time in other time slots increases with volume. The Highway Capacity Manual (2000) grades the quality or level of service (LOS) of transportation facilities on an A-F scale, with "A" being the best and "F" the worst. Based on that manual, the LOS of time slots minus 2, -1, 0, 1, 2 are A, C, D, C, and A, respectively.

Static Strategy

In this section, a fixed price is used for the HOT lane during the peak period (3-8 p.m.). Although a static pricing plan does not give the planner the ability to control the traffic or maximize revenue, this type of pricing saves costs with infrequent price changes. With dynamic tolls, the planner needs more equipment to control the toll and announce it to travelers. Static tolls help travelers choose their routes before approaching the entrance of HOT lane.

However, there are some disadvantages with fixed tolls. The inability to use price as a tool to control traffic or to maximize profit is significant. In addition, the constraint of having a certain level of service in the HOT lane makes the fixed toll policy less attractive: a low toll makes the HOT lane over-utilized and increases the possibilty that traffic in the middle time slots will rise to unacceptable levels, and a high toll encourages travelers to avoid the HOT lane in the shoulder time slots (i.e.,

time slots -2, 2), leading to their under-utilization. To illustrate, Table 2 provides a numerical study of two different toll levels ($1 and $2) over the peak period.

Table 2: Static Pricing Comparison

Road		Static 1				Static 2			
Type	Period	Flow	VOC[a]	TT	Speed	Flow	VOC	TT	Speed
	-2 (3-4)	2272	1.01	28.06	42.77	2632	1.17	35.76	33.56
Regular	-1 (4-5)	2594	1.15	34.78	34.50	2890	1.28	43.60	27.53
Lane	0 (5-6)	3295	1.46	60.90	19.70	3505	1.56	72.85	16.47
	1 (6-7)	2783	1.24	40.06	29.96	3045	1.35	49.44	24.27
	2 (7-8)	2388	1.06	30.18	39.76	2734	1.21	38.58	31.11
	-2 (3-4)	86	0.04	18.46	65.00	10	0.01	18.46	65.00
HOT	-1 (4-5)	1703	0.77	21.68	55.35	904	0.40	18.70	64.16
Lane	0 (5-6)	2706	1.22	38.93	30.82	2476	1.10	32.22	37.25
	1 (6-7)	2065	0.93	25.27	47.49	1782	0.80	22.27	53.89
	2 (7-8)	108	0.05	18.46	65.00	19	0.01	18.46	65.00

[a]If VOC ≥ 1, then the road is congested (indicated by shaded cells).

These two cases are Static 1 and Static 2, respectively. Compared to the Free-HOT-Lane case, fewer travelers use the HOT lane because it is no longer free. In both cases, the regular lane is over-utilized, and there is one case of overflow in time slot zero. The travel time of time slot zero is 60.9 minutes and causes 18 travelers to overflow into the next time slot. The time slots with overflow are highlighted with dark shadows.

Both cases result in the HOT lane not meeting its minimum level of service (VOC ≤ 0.9). Because the tolls are not high enough to prevent a large number of travelers from using time slots -1, 0, 1 the HOT lane is underutilized during the shoulder time slots and over-utilized in the middle time slots.

Compared with the Free-HOT-Lane case, the flows on the regular lane in Static 1 and Static 2 are large, with a VOC of 0.4 - 1.33 in the Free HOT lane (Table 1) and VOC from 1.17 - 1.56 in the Static 2 case (Table 2). On the other hand, the number of HOT lane travelers in time slots -2 and 2 drops significantly, with VOC near 0 in the shoulder time slots of Static 1 and Static 2 versus a VOC of 0.75 in the shoulder time slot of the Free-HOT-lane case.

A $2 toll results in some travelers leaving earlier or later to pay no toll and enjoy the better LOS in the regular lane during time slots -2 and 2. When Static 1 and Static 2 are compared, it is clear that those travelers who prefer to use the middle time slots in Static1 change their travel times in Static 2 and use the shoulder time slots. For instance, the difference between in VOC_{-2}^{NT} Static 2 and Static 1 is 0.16 while the difference between VOC_0^{NT} in Static 2 and Static 1 is 0.1, suggesting that higher tolls make travelers choose shoulder times. This fact implies that some travelers will alter their route choices and departure times if the right toll is charged.

Dynamic Strategy

The previous section showed that a fixed toll during the peak period does not efficiently allocate travelers among the time slots. When the shoulder time slots of the HOT lane are almost empty, a high fixed toll leads to the over-utilization of the middle time slots and violates the minimum LOS in a few cases. On the other hand, a low fixed toll causes the HOT lane to be over-utilized for most time slots. To cope with this problem, a dynamic policy is employed.

Table 3: Dynamic Pricing: Comparing Basic and Proposed Model

Road Type	Period	Dynamic 1 $p = [1, 2, 4, 2, 1]$				Dynamic 2 $p = [0.8, 1.7, 3.5, 2, ,0.6]$			
		Flow	VOC[a]	TT	Speed	Flow	VOC	TT	Speed
Regular Lane	-2 (3-4)	2546	1.13	33.60	35.71	2350	1.04	29.45	40.75
	-1 (4-5)	2813	1.25	41.02	29.25	2646	1.18	36.12	33.23
	0 (5-6)	3447	1.53	69.31	17.31	3072	1.37	50.54	23.74
	1 (6-7)	3158	1.33	47.17	25.44	2876	1.28	43.10	27.84
	2 (7-8)	2656	1.18	36.38	32.99	2450	1.09	31.44	38.17
HOT Lane	-2 (3-4)	898	0.40	18.70	64.19	1147	0.51	19.08	62.88
	-1 (4-5)	1043	0.46	18.89	63.53	1227	0.55	19.28	62.25
	0 (5-6)	1609	0.72	20.88	57.48	1839	0.82	22.58	53.13
	1 (6-7)	1373	0.61	19.74	60.78	1410	0.63	19.89	60.35
	2 (7-8)	626	0.28	18.52	64.81	983	0.44	18.80	63.84

[a]If VOC ≥ 1, then the road is congested (indicated by shaded cells).

To study the performance of the existing toll strategies, the toll menu in effect on Interstate 15 in San Diego, California, are used (I-15 FasTrak Program 2005). In this menu, the toll levels of time slots -2, -1, 0, 1, and 2 were fixed at $1, $2, $4, $2, and $1, respectively, or $p = [1, 2, 4, 2, 1]$. This is called Dynamic 1 in this study. Table 3 illustrates the results for two different toll menus (Dynamic 1 and Dynamic 2) where the first set of columns illustrates the results for Dynamic 1. Due to the high prices charged in all the time slots, many travelers use the regular lane, and the HOT lane is underutilized.

The VOC of the HOT lane ranges from 0.28 to 0.72, and, except for one time slot, the average speed is above 60 miles per hour, which is very close to the free flow speed. However, the regular lane is over-utilized, with VOC as high as 1.53 in time slot 0. Time slot 0 also has an overflow of 170 travelers because the travel time is greater than 60 minutes.

When $p = [0.8, 1.7, 3.5, 2, 0.6]$, hereafter Dynamic 2, the utilization of the HOT lane increases significantly and congestion on the regular lane is reduced. The VOC of the shoulder time slots of the HOT lane increases from 0.28 to 0.44, while the overall VOC of the HOT lane remains in the accepted range.

The traffic flow on the regular lane in the Dynamic 2 decreases by 1,056 travelers compared to Dynamic 1. More middle time slot travelers switch routes since VOC_0^{NT} deceases by 0.12, while VOC_{-2}^{NT} and VOC_2^{NT} decrease by 0.9. Many of these travelers also change their travel times. The 1,056 travelers who switch routes use time slots -1, 2, -2 instead of time slots minus 1, 0, 1. The VOC of slots -2, 2 increases by 0.11 and 0.16, respectively, compared to increases of 0.02 and 0.08, respectively, for time slots -1, 1. This shows that assessing an appropriate toll makes more travelers change their travel times and depart when there is less traffic. This results in less congestion on the regular lane and increases HOT lane utilization while maintaining the minimum level of service.

Table 4 compares the performace of the different pricing policies and summarizes other factors (i.e., total profit from toll collection, consumer generalized cost, average travel time of consumers, and total overflow). As expected, total profits in the No-HOT-Lane and Free-HOT-Lane cases are 0. Interestingly enough, although Dynamic 2 significantly reduces generalized cost and average travel time, the planner gains almost as much profit as using Dynamic 1. The average generalized cost for consumers is very high in the base model where one less lane is available for the travelers. But consumers enjoy the lowest generalized cost in the Free-HOT-Lane case because two free lanes are available.

Table 4: Comparing the Performances of Different Tolling Policies

Type	Profit	Consumer Cost	Avg. TT	$E(TT - FFTT)^2$	Overflow
NO HOT lane	0	47.55	113.75	132.08	3055
Free HOT lane	0	18.37	31.63	14.33	0
Static 1	6667.79	19.67	36.86	18.07	18
Static 2	10385.48	20.48	43.43	24.92	228
Dynamic 1	12794.80	27.17	39.31	22.22	170
Dynamic 2	12738.96	21.13	33.45	16.09	0

Another index in the table is $E(TT - FFTT)$,[2] which captures the expected squared difference between travel time and free flow travel time in each time slot. This index measures the overall performance of each tolling policy in terms of consumer travel time, and assigns a higher penalty to those time slots with high travel times. Among the cases with HOT lane, $E(TT - FFTT)^2$ is the highest in Static 2 since more travelers use the middle time slot on the regular lane. In comparing Dynamic 1 and Dynamic 2, $E(TT - FFTT)^2$ significantly reduces in the latter because the distribution of consumers among different time slots is flat and the time slots are congested.

CONCLUSIONS

The findings in this paper provide insight into congestion pricing and highway management. Upon arrival at a multilane highway where one lane is free and another one is subject to toll collection, a traveler observes the current toll and decides whether or not to enter the HOT lane. Some travelers are willing to change their departure times and or routes to face lower generalized costs. A user equilibrium method was used to determine the distribution of travelers between HOT and regular lanes during the peak period, and to compare the impacts of different pricing policies on traveler behavior. It is shown by numerical example that a planner can manage congestion levels with tolls. Further, a price menu currently in effect on Interstate 15 in San Diego, California, was used to show that the congestion level on the regular untolled lane can be reduced without loss of profit from toll collection. An extension of this study is to develop an optimization model to calculate a pricing menu in which departure rates are uniform during the peak period. A multi-objective optimization model that both reduces the regular lane's congestion level and maximizes the total expected profit can also be developed. Finally, one can define multiple classes of travelers who have different values of time and disutilities of changing their departure times.

Endnotes

1. For a comprehensive review, the readers are referred to Kutz (1992).

References

Arnott, R., A. de Palma, and R. Lindsey. "A Structural Model of Peak-period Congestion: A Traffic Bottleneck with Elastic Demand." *American Economic Review* 83 (1), (1993): 161-179.

Arnott, R., A. dePalma, and R. Lindsey. "Departure Time and Route Choice for the Morning Commute." *Transportation Research* B 24 (3), (1990): 209-228.

Braid, R. "Peak-load Pricing of a Transportation Route with an Unpriced Substitute." *Journal of Urban Economics* 40 (2), (1996): 179-197.

Carey, M. and A. Srinivasan. "Externalities, Average and Marginal Costs, and Tolls on Congested Network with Time-varying Flows." *Operations Research* 41 (1), (1993): 217-231.

Dafermos, S. and F. Sparrow. "Optimal Resource Allocation and Toll Patterns in User-optimized Transportation Network." *Journal of Transportation Economics and Policy* 5, (1971): 198-200.

Fielding, G. and D. Klein. "High Occupancy Toll Lanes: Phasing in Congestion Pricing a Lane at a Time." *Reason Foundation, Policy Study* 170, (1993).

Frank, M. and P. Wolfe. "An Algorithm for Quadratic Programming." *Naval Research Logistics Quarterly* 3 (1956): 95-110.

Highway Capacity Manual, *Transportation Research Board*, Washington, D.C., 2000.

I-15 FasTrak Program, Wilbur Smith Associates. "I-15 Express Lanes Toll Modification Study." December 2005. Available in sandag.org.

Jeihani, M., P. James, and A. Saka. "Measuring Non-recurring Post-incident Traffic Recovery Time: Comparing Shock Wave Theory and Simulation Modeling." *Maryland State Highway Administration,* Office of Traffic Safety, Hanover, MD, 2008.

Kumar, C. P., D. Basu, and B. Maitra. "Modeling Generalized Cost of Travel for Rural Bus Users: A Case Study." *Journal of Public Transportation* 7 (2), (2004): 59-72.

Kutz, M. *Handbook of Transportation Engineering.* McGraw-Hill Handbooks, 1992.

Litman, T. *Transportation Elasticities, How Prices and Other Factors Affect Travel Behavior.* Victoria Transportation Policy Institute, Victoria, B.C. Canada, 2008.

Liu, L. and J. McDonald. "Economic Efficiency of Second-best Congestion Pricing Schemes in Urban Highway Systems." *Transportation Research B* 33 (3), (1999): 157-188.

Oketch, T. and M. Carrick. "Calibration and Validation of a Micro-simulation Model in Network Analysis." *The Proceedings of the 84rd TRB Annual Meeting*, Washington, D.C., 2005.

Small, K. and J. Yan. "The Value of Value Pricing of Roads: Second-best Pricing and Product Differentiation." *Journal of Urban Economics* 49 (2), (2001): 310-336.

Verhoef, E., P. Nijkamp, and P. Rietveld. "Second-best Congestion Pricing: the Case of an Untolled Alternative." *Journal of Urban Economics* 40 (3), (1996): 279-302.

Vickrey, W. "Congestion Theory and Transportation Investment." *American Economic Review* 59,s (1969): 251-261.

Yang, H. and H. Huang. "Carpooling and Congestion Pricing in a Multilane Highway with High-occupancy-vehicle Lanes." *Transportation Research A* 33 (2), (1999): 139-155.

Yang, H. and Q. Meng. "Departure Time, Route Choice and Congestion Toll in a Queuing Network with Elastic Demand." *Transportation Research Part B* 32(4), (1998): 247-260.

Soheil Sibdari *holds the position of assistant professor in the Charlton College of Business, University of Massachusetts at Dartmouth. He received an M.S. in economics and a Ph.D. in industrial engineering from Virginia Tech. His research interests include dynamic pricing, decision making under uncertainty, transportation sciences, and game theory.*

Mansoureh Jeihani *is an assistant professor in the Department of Transportation and Urban Infrastructure Studies at Morgan State University. She received a Ph.D. in civil engineering (transportation systems) and an M.S. in economics from Virginia Tech, a master's degree in socio-economics systems engineering from IRPD, and a B.Sc. in computer engineering from Iran National University. Her research interests are in transportation planning, transportation economics, intelligent transportation systems, and traveler behavior.*