

Optimal Dynamic Pricing Strategies for High-Occupancy/Toll Lanes

Yingyan Lou

Graduate Research Assistant

Department of Civil and Coastal Engineering

University of Florida, Gainesville, FL 32611-6580

Yafeng Yin, Ph.D.*

Assistant Professor

Department of Civil and Coastal Engineering

University of Florida, Gainesville, FL 32611-6580

Jorge A. Laval, Ph.D.

Assistant Professor

School of Civil and Environmental Engineering

Georgia Institute of Technology, Atlanta, GA 30332-0355

* Corresponding Author. Tel. 352-392-9537, Email: yafeng@ce.ufl.edu

Optimal dynamic pricing strategies for high-occupancy/toll lanes

Yingyan Lou, Yafeng Yin and Jorge A. Laval

July 8, 2009

Abstract

This paper proposes a self-learning approach to determine optimal pricing strategies for high-occupancy/toll lane operations. The approach learns recursively motorists' willingness to pay by mining the loop detector data, and then specifies toll rates to maximize the freeway's throughput while ensuring a superior travel service to the users of the toll lanes. In determination of the tolls, a multi-lane hybrid traffic flow model is used to explicitly consider the impacts of the lane-changing behaviors before the entry points of the toll lanes on throughput and travel time. Simulation experiments are conducted to demonstrate and validate the proposed approach, and provide insights on when to convert high-occupancy lanes to toll lanes.

Keywords: high-occupancy/toll lanes, dynamic pricing, self-learning, lane changes

1. Introduction

Road pricing has been advocated as an efficient way to reduce congestion since the seminal work by Pigou (1920) and Knight (1924) (see Lindsey and Verhoef, 2001, and Lindsey, 2006, for recent reviews). However, it has only recently been adopted perhaps due to the advent of electronic tolling and the pressing need for alternative funding sources to finance transportation projects. For example, Singapore implemented its Area Licensing Scheme to restrict vehicular traffic into the city's central area in 1975. Later (1988) it was renamed Electronic Road Pricing, in part to reflect the use of new technology. In Norway, the first toll ring was operational in Bergen in 1986 and, subsequently, two additional toll rings were established in Oslo and Trondheim. More recently, the city of London introduced in February 2003 a five-pound (later increased to eight) daily fee on cars entering its city center. In the U.S., a more prevalent form of congestion pricing is high-occupancy/toll (HOT) lanes, which refer to high-occupancy vehicle (HOV) facilities that allow lower-occupancy vehicles to pay a toll to gain the access. Since the first HOT lane was implemented in 1995 on State Route 91 in Orange County, California, the concept has been becoming popular among governors and transportation officials, in state legislatures and the media (Orski, 2006).

Often time operation policies of HOT lanes are to provide a superior free-flow traffic service on the toll lanes while maximizing the throughput rate of the freeway, i.e., the combined throughput of both regular and toll lanes (FHWA, 2003). Between these two objectives, the operators often give higher priority to the former, because the HOV lanes are designed "first and foremost to provide less congested conditions for carpoolers and

transit users” (Munnich, 2006). To achieve the objectives efficiently, tolls should be adjusted real time in response to changes in traffic conditions. In practice, several transportation authorities price their toll lanes dynamically, although in an ad-hoc manner. For example, the toll rates for I-394 HOT lanes in Minnesota can be adjusted as often as every three minutes. When a change in the density occurs, the rate is adjusted upward or downward according to a pre-defined “look-up” table (Halvorson, et al., 2006). The literature does not offer a practical and sensible approach either. Previous studies (see, e.g., Arnott et al., 1998; Chu, 1995; Liu and McDonald, 1999; Yang and Huang, 1997 and Kuwahara, 2001) have examined time-varying tolls for bottlenecks. However, most, if not all, of these studies consider hypothetical and idealized situations in which analytical solutions can be derived. For example, the travel demand or the demand function is usually assumed known.

Recently, Yin and Lou (2009) delivered a proof of concept of a reactive self-learning approach for determining time-varying tolls in response to the detected traffic arrivals. The approach learns in a sequential fashion motorists’ willingness to pay and then determines pricing strategies based upon a point-queue model. This paper further develops this approach with a more realistic representation of traffic dynamics and an explicit formulation for toll optimization. The multi-lane hybrid traffic flow model proposed by Laval and Daganzo (2006) is incorporated to explicitly consider the impacts of the lane-changing behaviors before the entry points of the toll lanes on the freeway throughput and travel time. A nonlinear optimization model is formulated to determine an optimal toll for a rolling horizon to maximize the throughput of the corridor while ensuring the density of HOT lane is less than the critical density. Moreover, extensions to

the proposed approach with more realistic cell representations for different HOT lane slip ramp configurations and more efficient proactive pricing strategies are discussed. This paper also sheds further light on the discussions when HOV lanes should be converted to HOT lanes.

For the remainder, Section 2 introduces the concept of reactive self-learning and Section 3 encapsulates the multi-lane hybrid model in the toll optimization problem. Subsequently, Section 4 presents a simulation experiment to demonstrate the proposed approach and Section 5 discusses a couple of model extensions on representing different HOT lane slip ramp configurations and proactive self-learning pricing. Section 6 concludes the paper.

2. Reactive Self-Learning Approach

2.1 Basic concept

The essential idea of the self-learning approach is that in the operation, motorists' (revealed) willingness to pay can be gradually learned by mining the loop detector data, and the attained knowledge can then be applied to determine optimal tolls for achieving control objectives.

More specifically, given a particular toll rate, we assume in this paper the proportion of the motorists willing to pay to gain access to the toll lane can be formulated as an aggregate Logit model whose parameters are unknown. However, since the flow rates before and after the lane choice can be measured directly by loop detectors, and travel

times on the toll and regular lanes can be either estimated or directly measured (via additional toll-tag readers installed along the freeway), such revealed-preference information can be used to estimate recursively the parameters of the Logit model. Based on the estimated Logit model (can be viewed as the demand function for the toll lane), the optimal toll rate can be determined explicitly to achieve the operation objectives.

In summary, the self-learning approach decomposes toll determination into two consecutive steps: first to use previous revealed-preference information to learn motorists' willingness to pay, and then to determine the optimal toll rate based on the detected approaching flow rates, the calibrated willingness to pay and the estimated travel times.

Two sets of detectors are required to materialize the idea (see Figure 1). The first set of detectors is installed before the toll-tag reader to detect the approaching traffic flows, while the second set of detectors is installed after the reader to detect the flows on the HOV/HOT and regular (general-purpose) lanes, respectively. Without loss of generality and to facilitate the presentation of the essential idea, this paper uses a simple setting shown in Figure 1 where there are one HOV/HOT lane and one regular lane, and a bottleneck is activated downstream. Moreover, it is a single segment of HOT lane with one entry, and there is no on/off ramp in between.

In Figure 1, $\mu_T(t)$ and $\mu_R(t)$ represent the approaching flow rates on HOV and regular lanes during time interval t respectively while $\lambda_T(t)$ and $\lambda_R(t)$ are the flow rates on HOT and regular lanes after the lane choice.

2.2 Calibration of willingness to pay

Given a specific toll rate $\beta(t)$, we adopt a Logit model to reflect motorists' decision on whether to choose the HOT lane. Assuming homogeneous motorists with the same willingness to pay, the relationship between the approaching flow rates and the rates on HOT and regular lanes is as follows:

$$\frac{\lambda_T(t) - \mu_T(t)}{\mu_R(t)} = \frac{1}{1 + \exp(\alpha_1(c_T(t) - c_R(t)) + \alpha_2\beta(t) + \gamma)} \quad (1)$$

where $c_T(t)$ and $c_R(t)$ are the (average) travel times on the HOT and regular lanes at time interval t . In Equation (1), there are three parameters to be estimated, α_1 , α_2 and γ , where α_1 and α_2 indicate respectively the marginal effect of travel time and toll on motorists' utility, and γ encapsulates other factors affecting motorists' willingness to pay. Note that α_1/α_2 represents motorists' trade-off between time savings and tolls, i.e., the value of travel time. For other variables in Equation (1), $\mu_T(t)$, $\mu_R(t)$ and $\lambda_T(t)$ can be obtained directly from loop detectors, $c_T(t)$ and $c_R(t)$ can be directly measured or estimated using traffic flow models, and $\beta(t)$ is set by the operator. It should be pointed out that here we use volume splits $(\lambda_T(t) - \mu_T(t))/\mu_R(t)$ to approximate the probabilities of lane choices.

In real-time operation, a recursive least-squares technique or discrete Kalman filtering (Ljung, 1999) can be used to estimate the constant parameters, α_1 , α_2 and γ . To do so, Equation (1) can be reformulated as follows:

$$\ln\left(\frac{\mu_R(t)}{\lambda_T(t) - \mu_T(t)} - 1\right) = \alpha_1(c_T(t) - c_R(t)) + \alpha_2\beta(t) + \gamma \quad (2)$$

Let $y(t) = \ln(\mu_R(t)/(\lambda_T(t) - \mu_T(t)) - 1)$, $x(t) = (\alpha_1(t) \ \alpha_2(t) \ \gamma(t))^T$ and

$H(t) = (c_T(t) - c_R(t) \ \beta(t) \ 1)$. According to Equation (2), the system/observation

equation can be concisely written as follows:

$$y(t) = H(t) \cdot x(t) + e_y$$

where $y(t)$ is the observed system output and e_y is a random measurement error with a mean of zero and a known variance of σ_y (the value could vary with different types of sensors).

Applying discrete Kalman filtering technique to estimate the parameters, we have:

$$\begin{cases} \begin{pmatrix} \hat{\alpha}_1(t+1) \\ \hat{\alpha}_2(t+1) \\ \hat{\gamma}(t+1) \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_1(t) \\ \hat{\alpha}_2(t) \\ \hat{\gamma}(t) \end{pmatrix} + G(t) \left[\ln\left(\frac{\mu_R(t)}{\lambda_T(t) - \mu_T(t)} - 1\right) - (c_T(t) - c_R(t) \ \beta(t) \ 1) \begin{pmatrix} \hat{\alpha}_1(t) \\ \hat{\alpha}_2(t) \\ \hat{\gamma}(t) \end{pmatrix} \right] \\ G(t) = P(t-1) \begin{pmatrix} c_T(t) - c_R(t) \\ \beta(t) \\ 1 \end{pmatrix} \left[(c_T(t) - c_R(t) \ \beta(t) \ 1) P(t-1) \begin{pmatrix} c_T(t) - c_R(t) \\ \beta(t) \\ 1 \end{pmatrix} + \sigma_y \right]^{-1} \\ P(t) = [I - G(t)(c_T(t) - c_R(t) \ \beta(t) \ 1)] P(t-1) \end{cases} \quad (3)$$

where the variables with “ $\hat{\ }^{\wedge}$ ” are estimates. P is the expected covariance matrix of the estimation errors and G is the Kalman gain. With an initialization of $\alpha_1(0), \alpha_2(0), \gamma(0)$ and $P(0)$, Equation (3) can update estimates of α_1, α_2 and γ real time with newly-obtained information. Because Kalman filter represents a recursive solution to the least-square parametric estimation problem, the difference caused by the choice of initial conditions between the recursive and static least-square estimates obtained from the same set of observations will become insignificant as time evolves and more data are made available (Ljung, 1999).

3 Optimal Pricing Strategies

The calibrated motorists' willingness to pay at interval t provides a basis for determining a toll for interval $t+1$. Consistent with the prevailing operation policies of HOT lanes, we now attempt to specify toll rates to maximize the throughput of the corridor while ensuring a superior traffic condition on the toll lane. For a more realistic representation of traffic dynamics, we adopt the multi-lane hybrid traffic flow model proposed by Laval and Daganzo (2006), motivated by the postulate that the lane-changing behaviors before the entry points to the HOT lane may create voids in traffic streams and reduce the throughput of the lane. In Laval and Daganzo's multi-lane hybrid model, each lane is modeled as a separate kinematic wave (KW) stream interrupted by lane-changing particles that completely block the traffic, and the flow transfers are predicted by using the incremental-transfer (IT) principle (Daganzo et al., 1997) for multiple-lane KW problems, coupled with a one-parameter model for discretionary lane-changing demand. In this paper, discretionary lane changes due to the positive speed difference between lanes are not considered. Of interest are the lane changes made by the low-occupancy vehicles that want to pay to access the HOT lane. The lane-changing demand is given by Equation (1). If the demand plus the arrival flow on the HOV lane exceeds available capacity of the HOT lane, the IT principle is used to prorate that available capacity¹.

¹ In this case, the number of lane changes estimated from the loop detector data does not necessarily represent the lane-changing demand. Therefore, when the HOT lane is congested, the detected flows will not be use to update the estimates of motorists' willingness to pay.

To implement the hybrid model, all lanes are partitioned into small cells of length Δx , in addition to discretizing the time into time intervals Δt . See Figure 2 for a discretized freeway representation.

A rolling-horizon framework is used to optimize toll rates, as sketched in Figure 3. Within the optimization rolling horizon (M time intervals but not the entire operation duration), it is assumed that traffic arrivals remain constant as the one detected at time t , namely $\mu_T(t)$ and $\mu_R(t)$, and the toll rate β is constant as well. The flow rates at cell m and n represent the throughput of the facility. The density of cell l serves as a level-of-service measure for the toll lane because this cell tends to suffer the most due to the blocking effect of lane-changing vehicles. It is a conservative way to formulate the problem to ensure the overall level-of-service along the freeway. The control objectives can now be more specifically stated as to maximize the sum of the flow rates at the downstream bottleneck (cell m and n) while ensuring the average density of cell l does not exceed the critical density. As aforementioned, the flow rates and density are determined by the multi-lane hybrid model and Equation (1), which essentially function as a mapping ψ :

$$(q_m, q_n, k_l) = \psi(\alpha_1, \alpha_2, \gamma, \beta, \mu_T(t), \mu_R(t)) \quad (4)$$

where q_m and q_n are the flow vectors for cell m and n whose elements are flow rates at each time interval within the optimization rolling horizon, and k_l is the density vector for cell l .

Consequently, the toll optimization problem can be written as:

$$\begin{aligned}
& \max_{\beta} \left\{ \sum_{j=t+1}^{t+M} [q_m(j) + q_n(j)] + \theta \cdot \min \left[\tilde{k} - \frac{1}{M} \sum_{j=t+1}^{t+M} k_l(j), 0 \right] \right\} \\
& \text{s.t. } 0 \leq \beta \leq \beta_{\max} \\
& \text{Condition (4)}
\end{aligned} \tag{5}$$

where θ is a penalty parameter; \tilde{k} is the critical density of the HOT lane and β_{\max} is the maximum toll rate specified by the tolling authority. The second term is a penalty function for density excess. If θ is selected appropriately, this term should be equal to zero with the optimal solution, ensuring that the average density of cell l is not greater than the critical density. Mathematically, as long as θ is large enough, i.e., the penalty term is sufficiently severe, the optimal toll rate can be found to maximize the throughput and ensure a free-flow traffic condition on the toll lane. Practically, a large θ may make the penalized problem ill-conditioned and thus difficult to solve. One may gradually increase the value of θ and solve a sequence of problems, generating a sequence of solutions that converge to the optimal toll. However, such an iterative approach is impractical for online implementation. Therefore, numerical experiments should be performed in advance to select an appropriate value of θ , making a tradeoff among computational tractability and those two control objectives.

The above problem is solved at each tolling interval to determine the toll rate for the next tolling interval.

4. Simulation Study

4.1 Design of Simulation Experiments

In order to demonstrate and validate the proposed self-learning approach, we conduct simulation experiments. The developed simulation platform consists of three major components: controller, monitor and simulator. The controller implements the self-learning approach. The monitor serves as a surveillance system, collecting information at each interval including flow rates before and after lane choices, densities, and travel times. The simulator attempts to replicate the motorists' lane-choice behaviors. At each interval, based on the toll rate specified by the controller and the instantaneous travel times from the monitor, the simulator applies the Logit model with the true values of α_1 , α_2 and γ to compute the percentages of the motorists for choosing the HOT lane.

The simulation site is a freeway segment shown in Figure 4. It is assumed that each lane obeys a triangular fundamental diagram as well as the downstream bottleneck. The relevant parameters are reported in Figure 4 and the small triangle is for the lanes in the downstream bottleneck. Additionally, all lane-changing vehicles are assumed to have an acceleration rate of 12.22 ft/s^2 , i.e., it will take a vehicle 7.2 seconds to accelerate from zero speed to free-flow speed.

The simulation duration is equally divided into discrete time intervals of 0.6 seconds, and all lanes are partitioned into small cells of 0.01 mile. To be consistent with the practice, the toll rate varies every two minutes, and the rolling horizon for toll optimization is 10 minutes. The weighting factor θ in Equation (5) is set as one and the toll optimization problem is solved using the golden-section method. In the simulation, random arrivals are generated from a virtual source with an average rate of 2400 vph for the regular lane and 600 vph for the HOV lane. The true values of α_1 , α_2 and γ used in the simulator are 0.5, 1 and 0.2. $\alpha_1(0)$, $\alpha_2(0)$, $\gamma(0)$ and $P(0)$ are initialized as 1, 2, 0 and

an identity matrix respectively.

4.2 Performance of Self-Learning Controller

Figure 5 is a proof of the concept that motorists' willingness to pay can be gradually learned. The figure presents the estimates of parameters α_1 , α_2 and γ , which converge from the initial values to the true values within four minutes.

Figure 6 presents the optimal toll rates determined by the self-learning controller and Figures 7 and 8 report the resulting throughputs and densities. It can be seen that the controller is able to adjust the toll real time to maintain a high and stable throughput of HOT lane while preventing the lane from being congested. The average maximum density along the lane is 30.31 vpmpl, less than the critical density of 40 vpmpl. Note that occasionally the density of toll lane is greater than the critical density. One of the reasons is that the toll optimization formulation (Equation 5) only penalizes the case that the average density across the optimization rolling horizon exceeds the critical density instead of the density for each individual time interval. Moreover, due to the randomness of user's choice behaviors and traffic arrivals, the realized traffic condition may be different from what the model predicted. The average throughput is 1149.6 vphpl. Although only 64% of the capacity of the downstream bottleneck, it is twice of the HOV volume, i.e., the utilization of HOV lane is doubled. One of the reasons for the HOT lane throughput not up to the capacity of the downstream bottleneck is the lane changes before the entry to the HOT lane. The lane-changing vehicles act as moving bottlenecks while accelerating to the speed prevailing on the HOT lane. They will create gaps in flow in

front of them that propagate forward, and thus reduce the throughput substantially. In a more ideal situation, the maximum HOT lane throughput we can achieve is approximately 80% of the downstream bottleneck capacity, as discussed later in section 4.3. In this simulation experiment, additional loss of throughput is due to initial inaccurate estimates of motorists' willingness to pay. We may also observe the high toll rates at the beginning of the simulation for the same reason.

Figures 7 and 8 also present the case of opening the HOV lane free of charge after the toll gate. Compared with the HOT lane, such an operation leads to almost the same throughput (1184.4 vphpl), but causes the toll lane congested (average maximum density is 51.8 vpmp). In view of the fact that the performance of the self-learning controller may depend on the value of the penalty coefficient θ , additional numerical experiments are conducted for different values of θ . When $\theta = 0.5$, the two performance measures are 34.18 vpmp and 1111.8 vphpl respectively; when $\theta = 2$, they are 27.71 vpmp and 1108.2 vphpl respectively. As expected, the average maximum density decreases when the penalty coefficient increases. With a smaller value of θ , the density of the toll lane may be occasionally beyond the critical density. All these three values result in comparable throughputs.

4.3 Conversion of HOV to HOT

There is a heated debate on when and where to implement HOT, HOV and general purpose lanes in the transportation community. For example, Dahlgren (2002) suggested that an added HOT lane performs (in terms of reducing delay) as well as or better than an

HOV lane in all circumstances. The result is established under the assumption that a toll can be correctly set such that the HOT lane can be fully utilized but not congested. This is consistent with the prevailing wisdom that converting HOV lane to HOT lane will not worsen traffic condition, if charging the right price. However, our observation from the simulation experiments suggests that there is a threshold in terms of HOV inflow, beyond which the HOV lane actually performs better than the HOT lane.

Figure 9 presents the (maximal) freeway throughputs with HOT and HOV operations respectively under various HOV inflow rates. The throughputs are obtained under the conditions that motorists' willingness to pay is known and the inflows are uniform for the entire simulation duration with the initial densities of the two lanes as 0 and 100 vpmpl respectively. The maximum throughput of the HOT lane under optimal pricing control is around 1440 vph (total freeway throughput reported in the figure minus the regular lane throughput of 1800 vph), 80% of the downstream capacity. A higher value could be achieved if the free flow condition of HOV lane is sacrificed. However, since the moving-bottleneck effect always exists, HOT lane may never be fully utilized.

Therefore, due to the impacts of the lane changes before the entry points to the HOT lane, if HOV flows are high enough, it may not be wise to convert HOV to HOT. In our experiments with a simple representation of lane slip ramp configurations (See 5.1 for a more detailed discussion), the threshold is approximately 80% of the downstream capacity. Certainly, if we further consider the costs incurred by toll collection, the value will be even lower. Additional simulation study should be performed to determine the threshold for a specific HOV/HOT lane design.

5. Model Extensions

This section discusses a couple of possibilities for improving the self-learning approach to be more realistic and efficient.

5.1 HOT Lane Slip Ramp Configurations

It is known that the lane-changing behaviors before the entrance of the HOT lanes will be largely affected by the physical configuration of access and egress of the HOT lanes. Figure 10 presents three typical designs of HOT lane slip ramp (FHWA, 2003). The cell representation scheme previously presented is relatively simple, and may reflect the reduced narrow buffer-separated (no weave lane) design but with a strong assumption that all the lane changes will be made at one location right before the end of the access segment. More realistic cell representation may be used, but the modeling of lane changes will have to be updated accordingly. For example, Figure 11 shows a cell representation scheme for the barrier-separated design option. The additional cells between the regular and the HOT lane represent the entrance (can be viewed as an acceleration lane). The weaving area may not need to be modeled as of more interest is the effect of the lane-changing behaviors on the HOT lane rather than the regular lane. The demand for the HOT lane is still calculated based on Equation (1). However, the lane changes from the acceleration lane to the HOT lane will then be modeled as discretionary behaviors, using the approach proposed in Laval and Daganzo (2006). As a consequence, the objective function for toll optimization should be slightly modified to consider the

average density of those cells most directly affected by lane changes (see Figure 11).

Denote those cells as a set L , the objective function in model (5) becomes

$$\sum_{j=t+1}^{t+M} [q_m(j) + q_n(j)] + \theta \cdot \min \left[\tilde{k} - \frac{1}{M \cdot |L|} \sum_{j=t+1}^{t+M} \sum_{l \in L} k_l(j), 0 \right]$$

where $|\bullet|$ denotes the cardinality of a set.

A numerical experiment is conducted based on this more realistic cell representation and the corresponding lane-change modeling. Figure 12 reports the optimal toll rates, with which the HOT lane throughput achieves 1750.8 vph and the average density of the acceleration lane and the relevant HOT lane segment is 40.55 vpmpl. Notice that such a high throughput is due to the fact that vehicles are able to accelerate before they merge into the HOT lane, and thus the voids created in the HOT lane flow are significantly reduced. These results suggest that HOT lane slip ramp configurations have substantial impacts on the performance of HOT lanes and the proposed modeling approach is able to capture those impacts.

5.2 Proactive Self-Learning Approach

The self-learning approach presented above determines a toll rate for each time interval in response to the inflows measured at that particular time. The tolls may thus fluctuate dramatically, which may cause safety issues in reality. If a driver who decides to access the toll lane observes a sudden price jump, he or she may become reluctant and make abrupt maneuvers that possibly lead to a rear-end accident. A proactive approach may be adopted to address the issue. Using historical data, short-term future inflows in

the rolling horizon can be predicted and the predication provides some lead time for adjusting toll rates in a smoother manner to respond the changes in traffic demand and maintain a high throughput. The problem (5) is revised to determine a vector of toll rates, β , for multiple future intervals within the rolling horizon.

$$\begin{aligned} \max_{\beta} & \left\{ \sum_{j=t+1}^{t+M} [q_m(j) + q_n(j)] + \theta \cdot \min \left[\tilde{k} - \frac{1}{M \cdot |L|} \sum_{j=t+1}^{t+M} \sum_{l \in L} k_l(j), 0 \right] \right\} \\ \text{s.t.} & \quad 0 \leq \beta_i \leq \beta_{\max} \quad \forall i = 1, 2, \dots, M \\ & \quad -\Delta\beta \leq \beta_i - \beta_{i-1} \leq \Delta\beta \quad \forall i = 1, 2, \dots, M \\ & \quad (q_m, q_n, k_l) = \psi(\alpha_1, \alpha_2, \gamma, \beta, \mu_T, \mu_R) \end{aligned}$$

In the above formulation, β_i denotes the toll rate of the i th interval within the rolling horizon and β_0 is the toll rate of the previous time interval, $\Delta\beta$ is the maximum toll variation for two consecutive intervals (due to safety consideration), and μ_T and μ_R are the corresponding vectors of future arrivals. Note that only the toll rate for the next interval will be applied in the operation because the optimization problem is solved every tolling interval to determine the toll vector based on the realized traffic condition and updated demand forecast.

To illustrate this proactive concept, a numerical example is conducted upon the simplest cell representation. An increase of the average inflow from 300 vph to 600 vph is arbitrarily created at time interval 10. The prediction is assumed to be 100% accurate, i.e. the actual future inflows are assumed to be known. Two toll schemes are used: Scheme A is determined interval by interval in response to the current detected inflow while in Scheme B, a toll vector for the next five intervals is computed each time based on the predicted future inflows and subject to a constraint that the variation of the tolls for two successive intervals is less than 0.25. Figure 13 compares these two schemes from

the simulation experiment. It can be observed that Scheme B changes in a smoother manner. The throughputs of HOT lane achieved by these two schemes are 1018.8 and 1075.8 vph respectively while the densities of cell l are 16.98 and 33.86 vpmpl respectively.

6. Concluding Remarks

This paper has developed and demonstrated a self-learning approach to determine dynamic pricing strategies for HOT lanes to provide superior travel services as well as efficiently utilize the capacities. The approach makes use of the data available from the loop detectors that are typically installed for HOT lane operations, and is thus cost-effective to implement.

A Logit lane choice model has been adopted in this paper to facilitate the presentation of the self-learning framework. Such a simple Logit model may not be very realistic in representing users' lane choices. However, the self-learning framework is very general and is able to many other choice models, including mixed Logit models with random coefficients that may be used to capture the heterogeneity. For example, Small et al. (2005) estimated a mixed Logit model to uncover the distribution of motorists' preference for travel time and reliability and Liu et al. (2007) calibrated a time-dependent mixed Logit model to estimate the time dependency of values of travel time and reliability using loop detector data from the California State Route 91. Those mixed-Logit models can be incorporated into the self-learning framework with the following considerations: 1) the calibration procedures based on Monte Carlo or quasi-Monte Carlo

simulation for mixed-Logit models are computationally demanding and thus not appropriate for on-line implementation. A hybrid of on-line and off-line estimations may be applied. The time-dependent parameters of the mixed Logit model estimated off-line serve as a regular pattern and then deviations from these parameters are assumed to be auto-regressive, estimated on-line using a Kalman filter; 2) Robust optimization techniques (e.g., Ben-Tal et al., 2007) can be applied to address proactively the heterogeneity in toll optimization. More specifically, we may seek robust pricing strategies that make the system performance less sensitive to variations of the calibrated coefficients.

The proposed approach is localized in that it determines a toll price for each segment or each entry point without any coordination among them. Localized control may not lead to a system optimum for the whole corridor. More importantly, such uncoordinated approaches may create inequality among users entering the managed lane from different entry points. For example, motorists who enter the toll lane via a downstream entry point may need to pay much higher toll for less amount of time saving. The inequality become severe when the time-saving per unit amount of the toll paid is significantly different among these motorists. Future research should be conducted to address this issue in addition to future exploration of the proactive pricing discussed in Section 5.2.

Acknowledgement

The paper benefits from our discussions with Prof. Siriphong Lawphongpanich at University of Florida and Prof. Yinhai Wang at University of Washington. The authors would like to thank the anonymous reviewers for their constructive comments that lead to

an improved presentation of our results. This research is supported in parts by grants from Florida Department of Transportation and the Center of Multimodal Solutions for Congestion Mitigation, University of Florida.

Reference

- Arnott, R., de Palma, A., and Lindsey, R. (1998). "Recent developments in the bottleneck model." *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility* (Kenneth J. Button and Erik T. Verhoef, eds), 79-110.
- Ben-Tal, A., El Ghaoui, L. and Nemirovski, A., (2006). "Forward: special issue on robust optimization." *Mathematical Programming, Series B*, 107, 1-3.
- Chu, X. (1995). "Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach." *Journal of Urban Economics*, 37, 324-343.
- Daganzo, C.F., Lin, W. And Del Castillo, J. (1997). "A simple physical principle for the simulation of freeways with special lanes and priority vehicles." *Transportation Research, Part B*, 31, 105-125.
- Dahlgren, J. (2002). "High-occupancy/toll lanes: where should they be implemented." *Transportation Research, Part A*, 36, 239-255.
- Federal Highway Administration (2003). *A Guide for HOT Lane Development*. U.S. Department of Transportation.
- Halvorson, R., Nookala, M. and Buckeye, K.R. (2006). "High occupancy toll lane innovations: I-394 MnPASS". *The 85th Annual Meeting of the Transportation Research Board, Compendium of Papers CD-ROM, No. 06-1265, January 9 – 13, 2006*.

- Knight, F.H. (1924). "Some fallacies in the interpretation of social cost." *Quarterly Journal of Economics*, 38, 582-606.
- Kuwahara, M. (2001). "A theoretical analysis on dynamic marginal cost pricing." *Proceedings of the Sixth Conference of Hong Kong Society for Transportation Studies*, 28-39.
- Laval, J.A. and Daganzo, C.F. (2006). "Lane-changing in traffic streams." *Transportation Research, Part B*, 40, 251-264.
- Lindsey, R. (2006). "Do economists reach a conclusion on road pricing? The intellectual history of an idea." *Econ Journal Watch*, 3, 292-379.
- Lindsey, R. and Verhoef, E. (2001). "Traffic congestion and congestion pricing." In *Handbook of Transport Systems and Traffic Control* (B.J. Button and D.A. Hensher ed.) Pergamon, 77-105.
- Liu, H.X., He, X. and Recker, W., (2007). "Estimation of the time-dependency of values of travel time and its reliability from loop detector data." *Transportation Research, Part B*, 41, 448-461
- Liu, L.N., and McDonald, J.F. (1999). "Economic efficiency of second-best congestion pricing schemes in urban highway systems." *Transportation Research, Part B*, 33, 157-188.
- Ljung, L. (1999) *System Identification: Theory for the User*, Prentice-Hall Inc., Upper Saddle River, NJ
- Munnich, L.W. (2006). "Easing the commute: Analysis shows MnPASS system curbing congestion on Interstate 394." *Downtown Journal/Skyway News*,

<<http://www.downtownjournal.com/articles/2006/07/24/news/news04.txt>> (July 24, 2006)

Orski, C. K. (2006). "Highway tolling has reached the tipping point." *Innovations Briefs*, 17.

Small, K.A., Winston, C. and Yan, J., 2005. "Uncovering the distribution of motorists' preferences for travel time and reliability." *Econometrica*, 73, 1367-1382.

Pigou, A.C. (1920). *Wealth and Welfare*, Macmillan, London.

Yang, H., and Huang, H.J. (1997). "Analysis of the time-varying pricing of a bottleneck with elastic demand using optimal control theory." *Transportation Research, Part B*, 31, 425-440.

Yin, Y. and Lou, Y. (2009). "Dynamic tolling strategies for managed lanes." *ASCE Journal of Transportation Engineering, Vol. 135, No. 2, 45-52.*

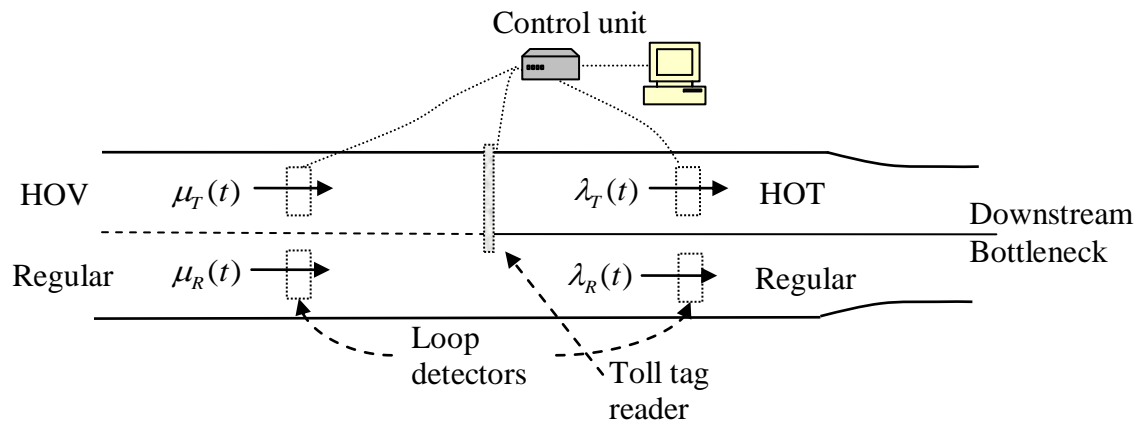


Figure 1: System configuration for the self-learning approach

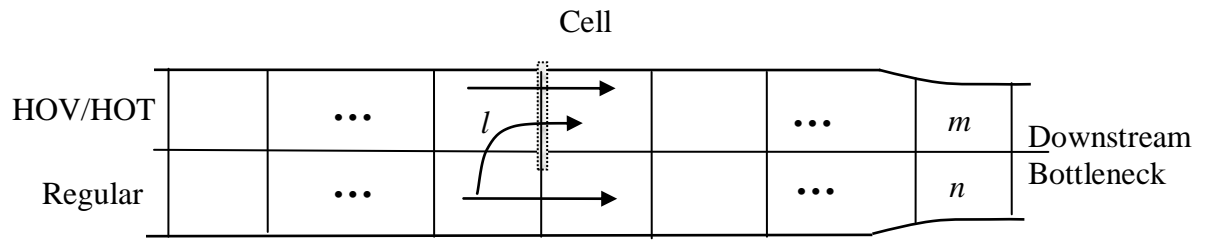


Figure 2: Discretized freeway representation

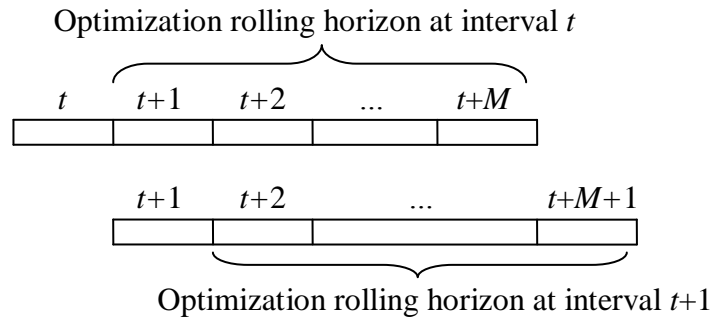


Figure 3: Rolling horizon scheme

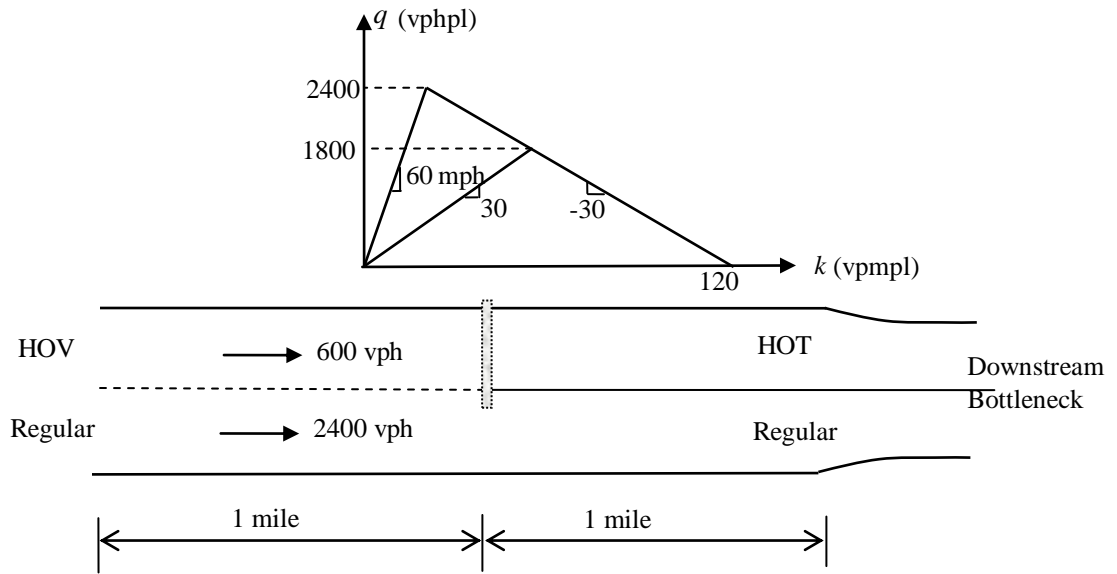


Figure 4: Simulation settings

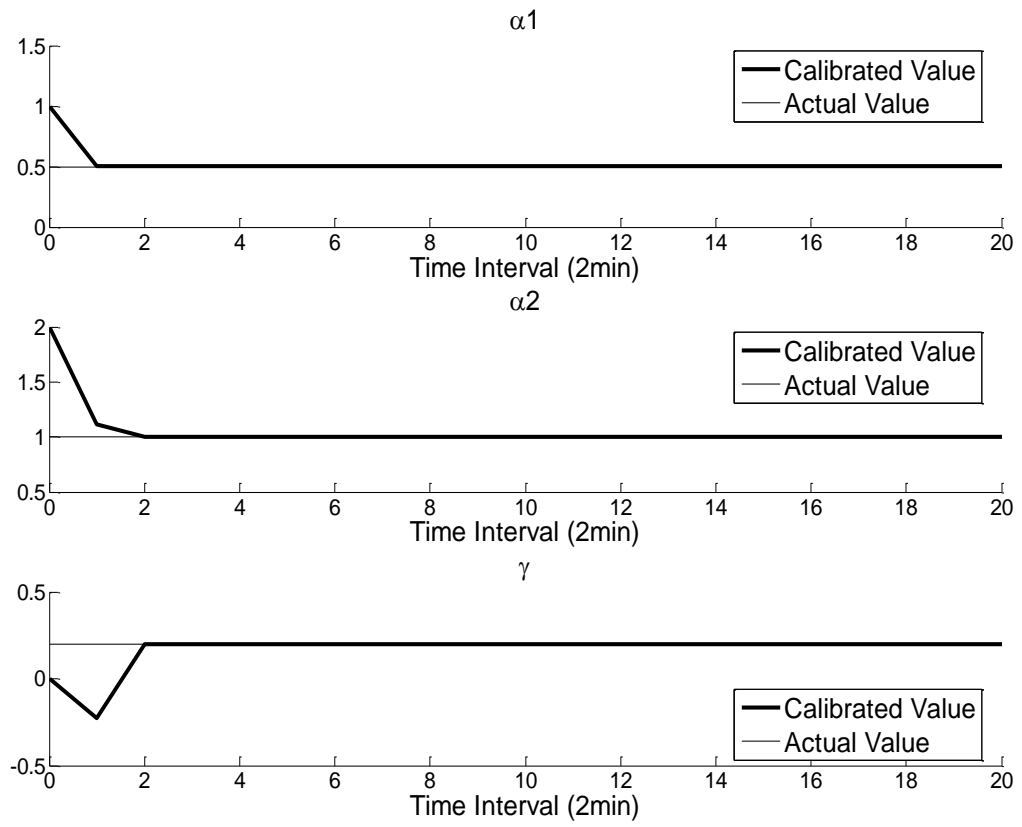


Figure 5: Calibrated α_1 , α_2 and γ

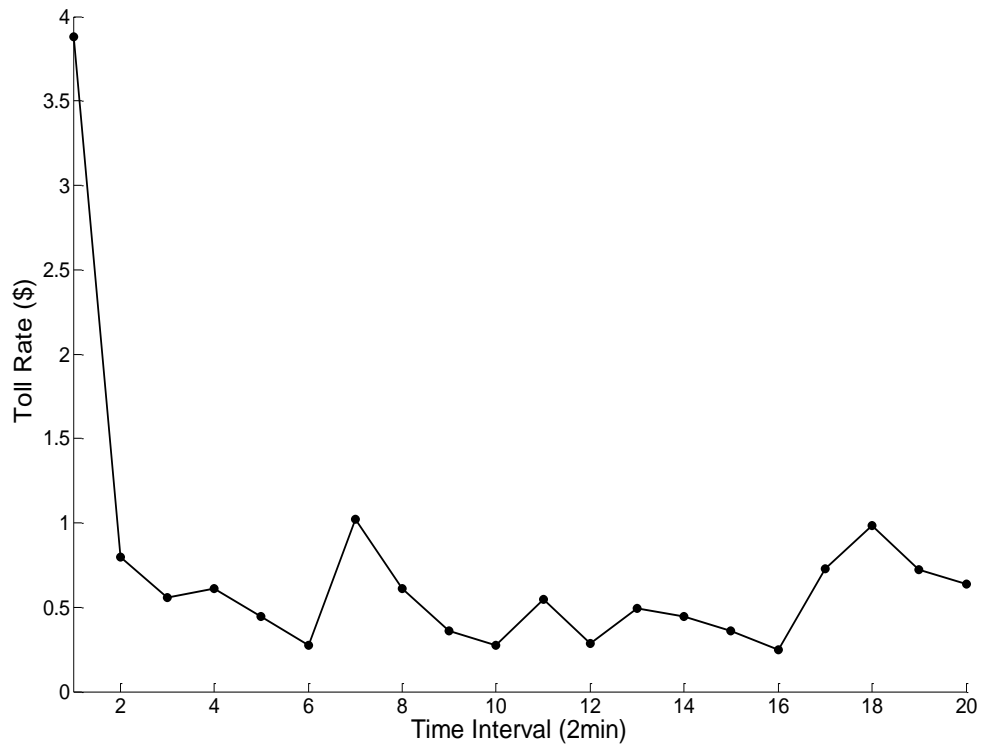


Figure 6: Optimal toll rates

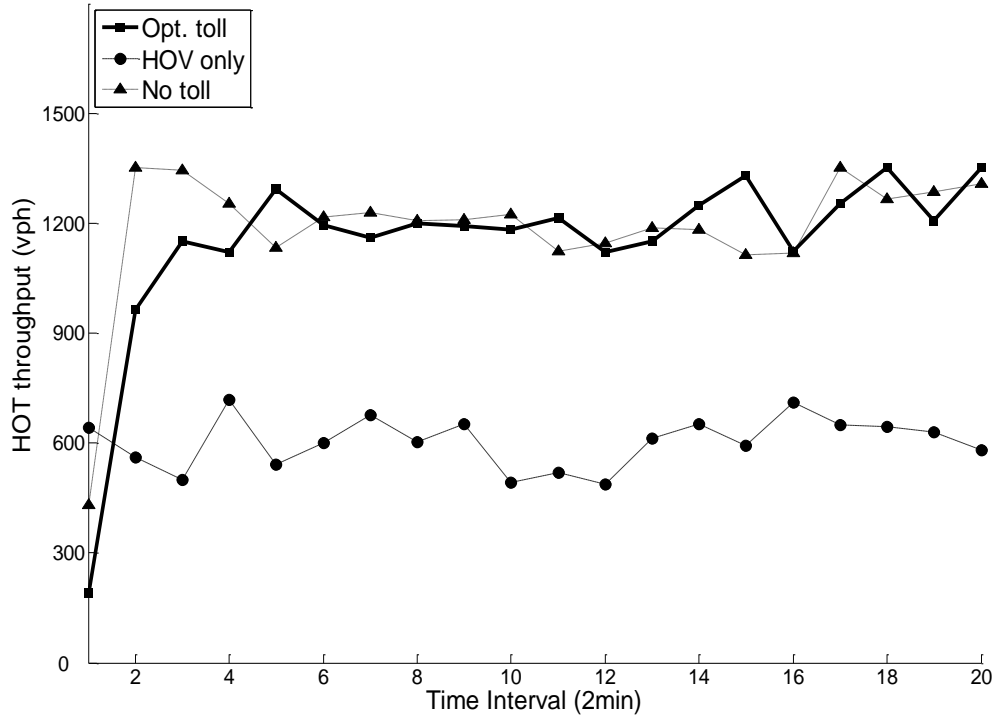


Figure 7: Resulting HOT throughputs

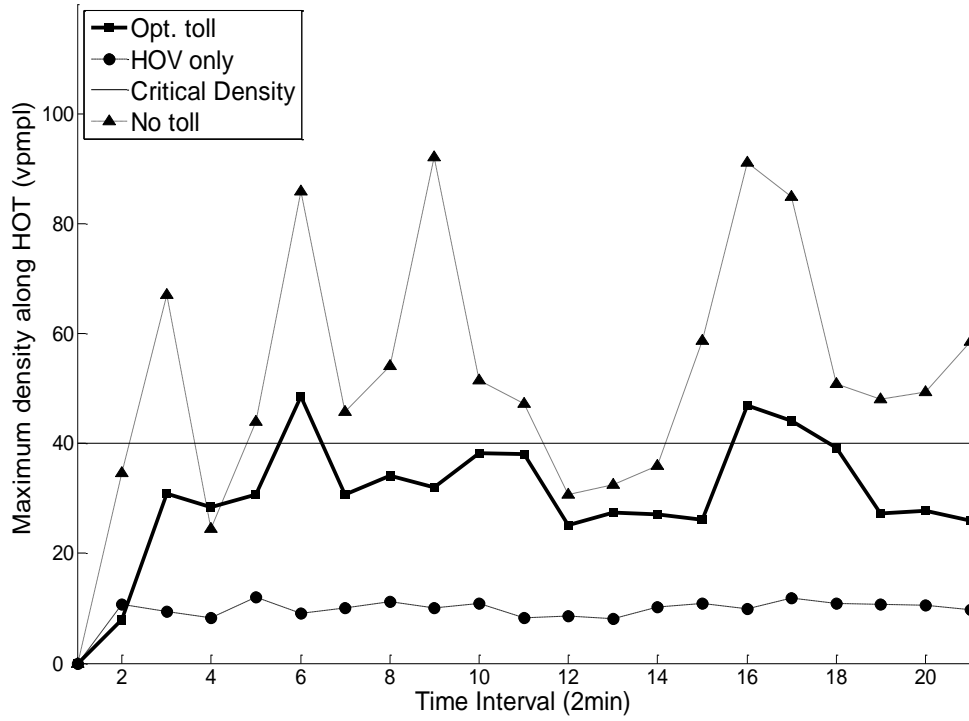


Figure 8: Resulting maximum density along the HOT lane

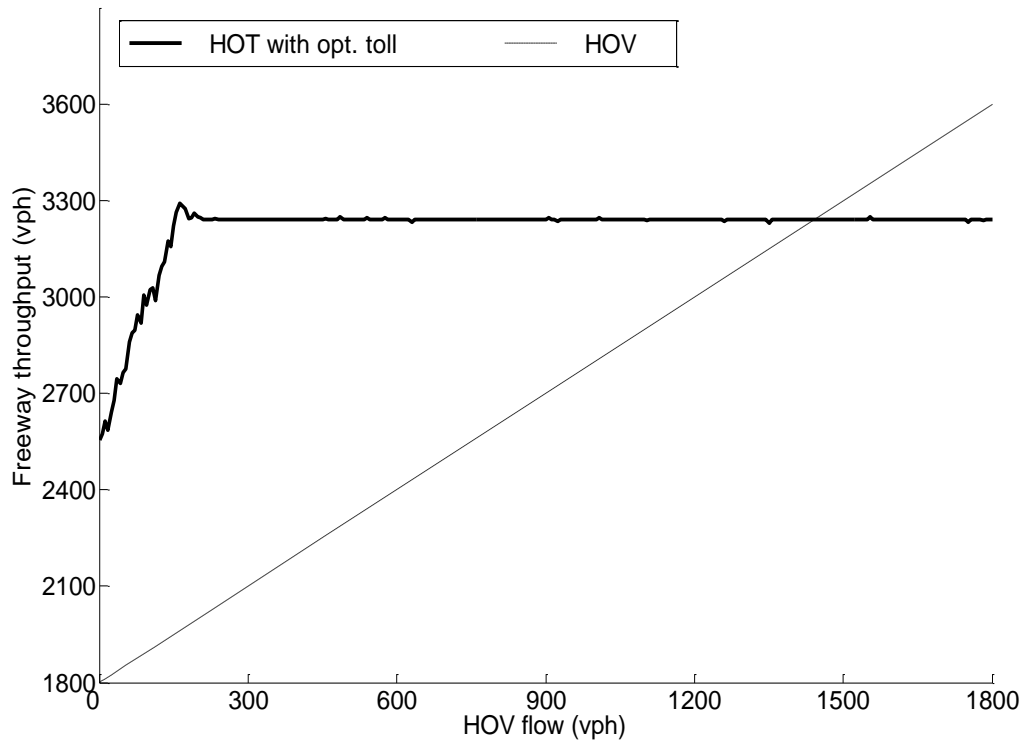


Figure 9: The freeway throughputs under HOV and HOT operations

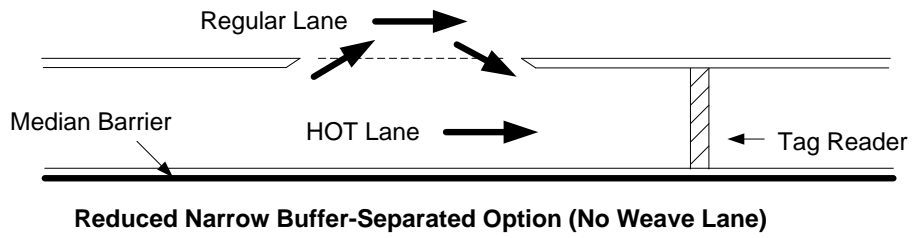
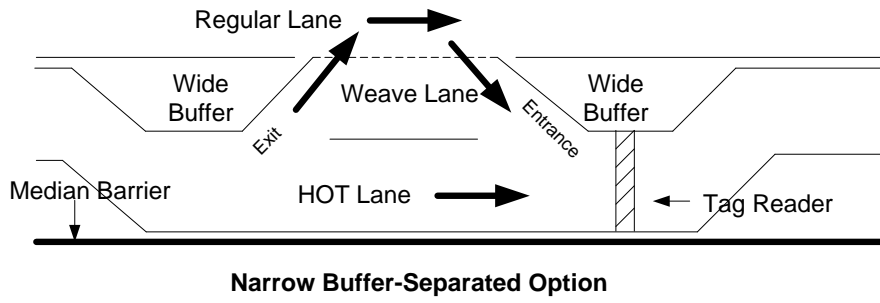
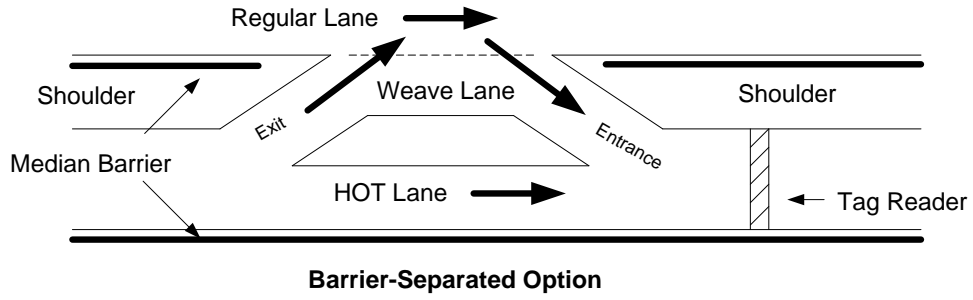


Figure 10: Alternative HOT lane slip ramp configurations

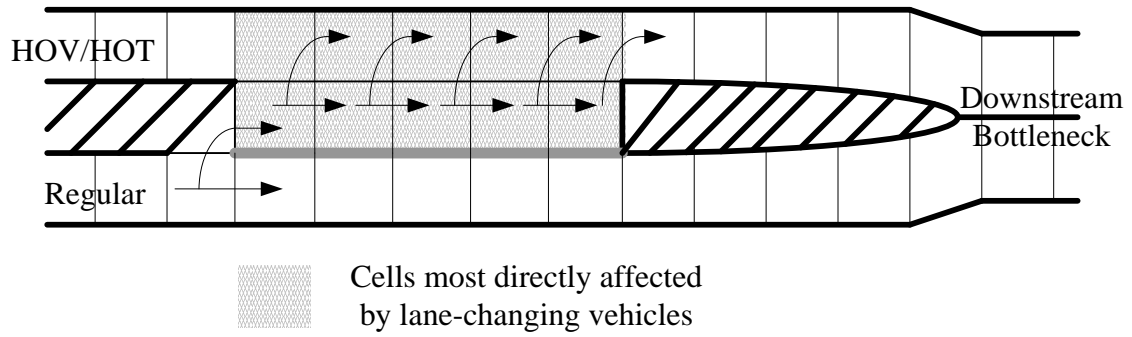


Figure 11: Discretized freeway representation (barrier-separated)

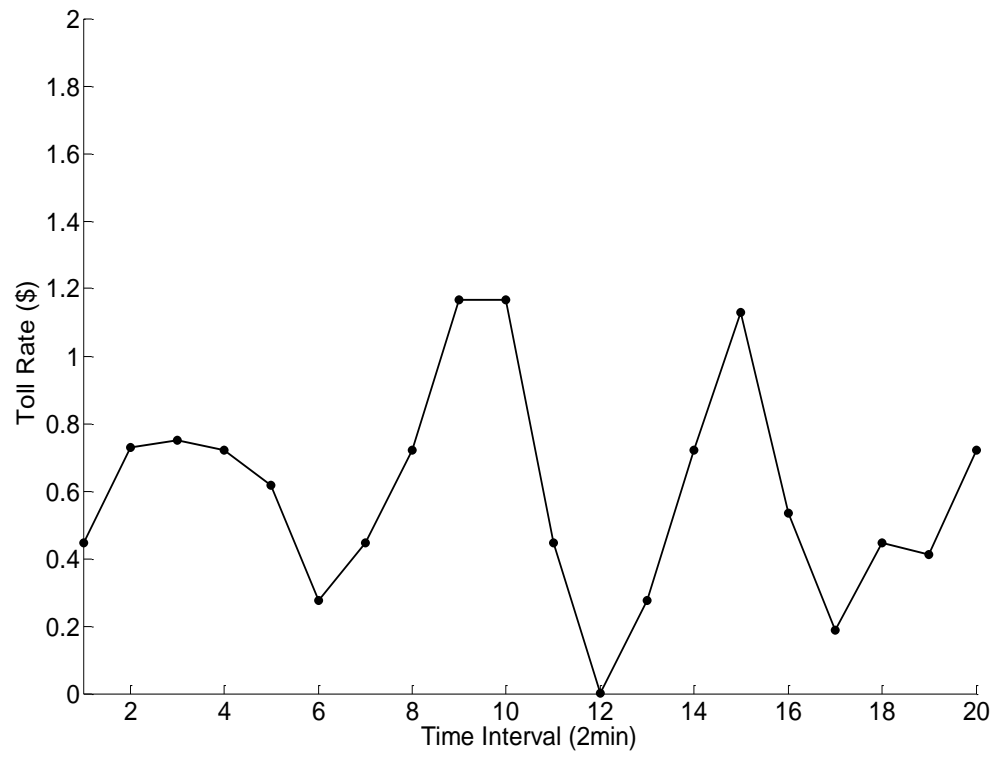


Figure 12: Optimal toll rates (barrier-separated freeway representation)

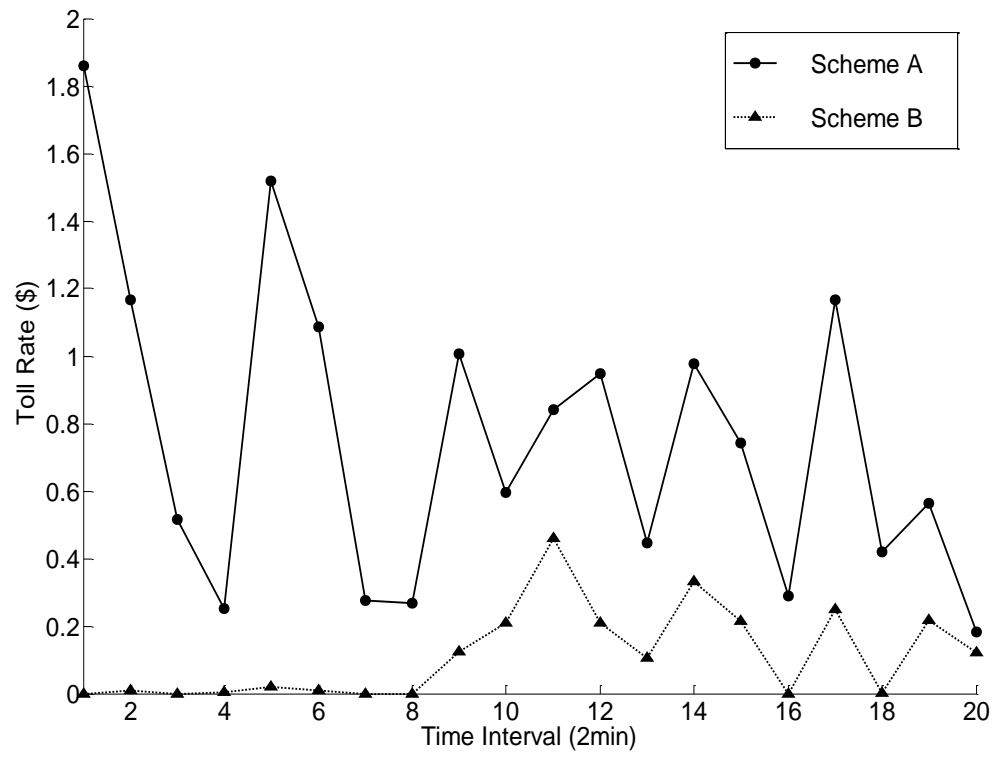


Figure 13: Reactive vs proactive pricing strategies