

2007

An Investigation of Congestion Pricing Options for Southbound Freight at the Pacific Highway Crossing

Matthew Roelofs
Western Washington University

Mark (Mark Christopher) Springer
Western Washington University

Follow this and additional works at: https://cedar.wvu.edu/bpri_publications



Part of the [Economics Commons](#), [Geography Commons](#), [International and Area Studies Commons](#), and the [International Relations Commons](#)

Recommended Citation

Roelofs, Matthew and Springer, Mark (Mark Christopher), "An Investigation of Congestion Pricing Options for Southbound Freight at the Pacific Highway Crossing" (2007). *Border Policy Research Institute Publications*. 85.

https://cedar.wvu.edu/bpri_publications/85

This Research Report is brought to you for free and open access by the Border Policy Research Institute at Western CEDAR. It has been accepted for inclusion in Border Policy Research Institute Publications by an authorized administrator of Western CEDAR. For more information, please contact westerncedar@wwu.edu.

An Investigation of Congestion Pricing Options for Southbound Freight at the Pacific Highway Crossing

Matthew R. Roelofs, Ph.D.
Associate Professor
Department of Economics
College of Business and Economics
Western Washington University

and

Mark Springer, Ph.D.
Associate Professor
Department of Decision Sciences
College of Business and Economics
Western Washington University

Research Report No. 4
August 2007

Border Policy Research Institute
Western Washington University
Bellingham, Washington
www.ac.wvu.edu/~bpri/

About the Border Policy Research Institute

The BPRI focuses on research that informs policy-makers on matters related to the Canada—U.S. border. Policy areas of importance include transportation and mobility, security, immigration, energy, environment, economics, and trade.

Border Policy Research Institute
Western Washington University
516 High Street
Bellingham, WA 98225-9110
(360) 650-3728

The BPRI encourages use of this report. Any part of the material may be duplicated with proper acknowledgment.

This research was supported by the U.S. Department of Transportation, Office of the Secretary, Grant No. DTOS59-05-G-00016.

This report is available at <http://www.ac.wvu.edu/~bpri/resources.html>

INTRODUCTION

Recent years have witnessed an increase in border security as well as continued growth in international truck traffic at the Pacific Highway Crossing (PHC) in Blaine, Washington. As noted in a recent study, the number of commercial vehicles crossing the border between British Columbia and Whatcom County, Washington, nearly doubled during the nineties, and nearly three-fourths of all trucks crossing this stretch of border are processed at PHC [1]. Congestion at PHC continues to be a concern; the study mentioned above found the average waiting time per southbound truck over a four-day period in the summer of 2002 to be about half an hour. Recent facility enhancements at PHC, notably additional capacity in the form of dedicated lanes and inspection booths for FAST (Free and Secure Trade) participants, were aimed at reducing this congestion. Other security-oriented changes, however, such as lengthier questioning at the inspection booths, have nonetheless increased overall processing times and doubled the waiting times for many trucks at the border [2].

Attempts have been made to quantify the cost of current and future waiting to the carriers and their supply chain partners. These studies frequently assume a flat hourly rate; the aforementioned study based the bulk of its economic projections on a 1995 Canadian study that valued driver time at forty-five dollars per hour (in then-current US dollars) for heavy trucks. Intuitively, however, we know that not all carriers place the same value on time throughout the day. The value a driver and carrier place on time depends on the opportunity cost of operating the vehicle, which may be affected by – among other things – what the driver is hauling and the schedule demands of the customer. In the case of limited border

processing capacity and a given number of trucks arriving at the border, U.S. Customs and Border Protection (CBP) could maximize total value across all carriers by more quickly processing those trucks for which time has a greater value.

By what means, however, could the CBP quickly and accurately identify those trucks that are most likely to benefit from a shorter wait time? A “congestion pricing” option for one of the border freight lanes would enable those carriers for whom time was most critical to “purchase” time. This would not mandate any slackening of the inspection process; buyers’ travel times would decrease because they would have access to a shorter line in the “pay” lane, not because they were receiving expedited

by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

treatment from the CBP. To ensure that total carrier value is kept high, the price for joining the lane would change depending on the relative time to be saved by joining it: a long “free” line, a short “pay” line, or a combination of both would increase the price of the “pay” lane, while the opposite line behavior would lower the price.

This study investigates the possibilities of providing a congestion pricing option for freight crossing from Canada into the U.S. We first review current theory and practice with regards to congestion pricing for transportation facilities. Next, we introduce the “test case” for our study: the southbound Pacific Highway Crossing for freight. We provide an overview of current PHC operations and discuss the simulation model and methodology used to investigate two alternatives for implementation of congestion pricing at PHC. The following section summarizes the results of our experiments, investigating the costs, benefits, and feasibility of introducing some form of congestion pricing at the commercial PHC. Finally, we conclude with a set of general policy recommendations.

**CONGESTION PRICING
FOR TRANSPORTATION:
CURRENT PRACTICE**

Most existing applications of congestion pricing in transportation are with

regards to roadways and, to a lesser extent, bridges. Charging motorists a toll for the use of transportation facilities is not new; the Pennsylvania Turnpike has been operating since 1940. In this traditional pricing approach, all users driving the same type of vehicle for the same distance pay the same amount, regardless of the road conditions or time of day. The primary objective of this non-variable pricing strategy is simply raising revenue. In contrast, a variable or congestion-based pricing strategy aims not solely to raise revenue, but to also achieve more effective use of transportation facilities by “managing demand” and “rationing supply.” In such a strategy, tolls are designed to vary along with demand for the facilities; due to these price differentials, some motorists are therefore encouraged to “time-shift” their travel demand to non-peak hours, and scarce rush hour capacity is effectively allocated to those motorists who place the greatest value on travel.

Since traffic demand patterns are somewhat predictable, one approach to congestion pricing has been to determine a time-based schedule, in advance, which specifies how rates change throughout the day. The State Route 91 (SR91) Express Lanes in Orange County, California, are a good and relatively sophisticated example of this approach. Tolls along this ten-mile stretch of road have a published schedule which is periodically updated, but generally remains

*by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.*

unchanged for at least six months. One advantage of this approach for users is price predictability. However, a potential downside is that road conditions may not correlate to the published schedule. Unpredicted changes in demand and roadway capacity can lead to a congestion pattern that is disconnected from the pricing schedule. Nevertheless, most congestion-pricing schemes for transportation facilities are time-based.

Dynamic pricing attempts to close the feedback loop between demand and price; prices are frequently updated, based on user demand, to maintain a specified level of service. California's I-15 Value Pricing Project, in which underutilized HOV (high-occupancy vehicle) lanes were opened to single-occupancy vehicles (SOV) for a congestion-based fee, was an early implementation of dynamic pricing. Traffic counts in the newly-renamed "HOT" (high occupancy/toll) lanes on I-15 were calculated for every six minute interval, and per-trip prices were adjusted based on this demand information; prices were adjusted to maintain a "level of service" (LOS) of C or better; which implies busy but relatively free flowing traffic. A toll schedule is published for the I-15 HOT lanes, but all of the tolls listed are simply estimates, and HOT lane management may override the published maximum toll.

It should be emphasized that congestion pricing on I-15 also differs from that on SR91 in that the I-15 express

lanes contain a mix of users: HOV carpoolers and toll-paying (SOV) drivers. A key element to making the I-15 HOT lanes successful was therefore the existence of excess capacity in the pre-conversion HOV lanes; also, maintaining free-flowing traffic on the new HOT lanes is politically necessary not only to give paying SOV drivers a sense of "value," but also to continue to encourage carpooling behavior by the public.

CONGESTION PRICING FOR TRANSPORTATION: CURRENT THEORY

The literature related to congestion pricing is voluminous and includes contributions by economists, engineers, and operations researchers. The theme that runs through all of this work is that when users of a transportation network impose costs on one another, private consumption of the transportation good will deviate from the social optimum, with a higher-than-optimal quantity consumed (i.e., too much traffic). The goal of any approach to pricing these goods is to force individuals to internalize the external cost they impose on other users and thus drive the interaction towards the social optimum. Note that this does not mean there will be no congestion, but rather an 'optimal' amount of congestion, an amount of traffic/congestion that correctly and completely balances the internal and

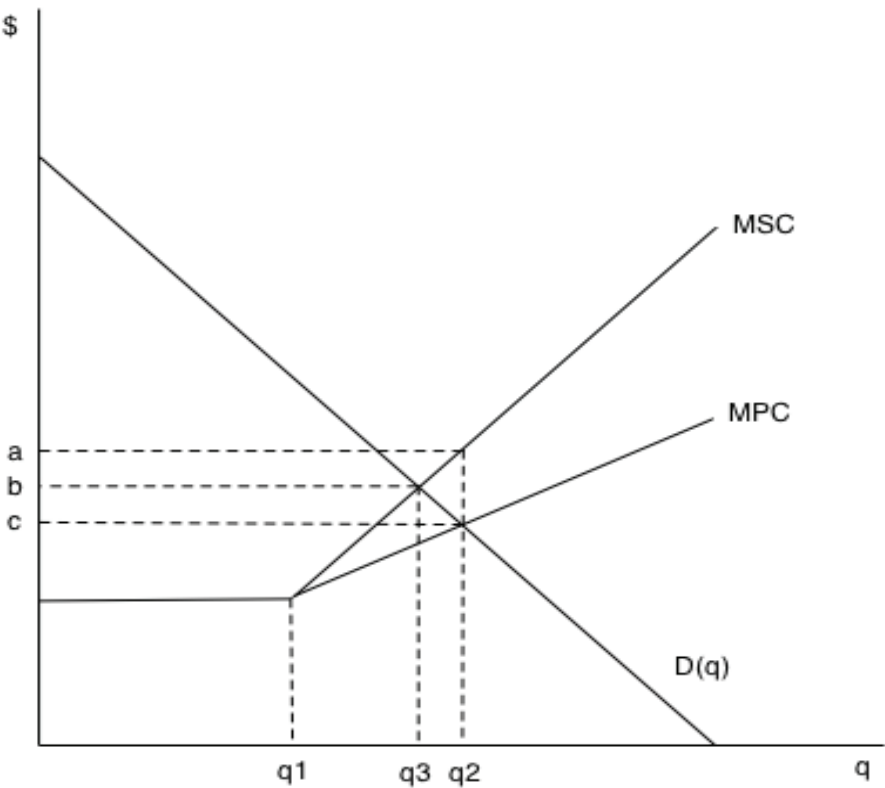
by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

external costs with the benefits of using that particular network at a particular time.

In Figure 1 below, we consider a simple theoretical scenario of a single link roadway – one origin, one destination, one path. We assume that the road users are homogeneous and have a combined inverse demand for travel given by $D(q)$, where q is the amount of transportation good consumed. In this simple example, q would be the number of trips made along the roadway per day.

The demand function derives from the value each user places on making the trip; as the “price” of making the trip increases, the number of trips demanded declines. In addition, each user incurs costs associated with the trip that include the obvious items such as fuel and vehicle depreciation as well as the dollar value of the time used to make the trip. These costs are assumed to be constant and identical for all users up to the point where the road becomes congested.

Figure 1. Optimal Toll for Single Link Roadway



by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

Once congestion begins to slow drivers down, both the time cost and the vehicle costs begin to increase as more users crowd the roadway. In addition, the portion of these costs borne by an individual driver and the total cost of that driver using the road begin to diverge. We call the costs paid by the driver her marginal private cost (MPC) and the full cost imposed – private cost plus any external costs imposed on other drivers – the marginal social cost (MSC). An alternative interpretation is to note that a given driver pays only the average cost of her road use and not the full marginal cost.

A quantity of road use between 0 and q_1 corresponds to the case where the roadway still has excess capacity. Once traffic levels rise above q_1 the roadway becomes congested and the MPC and MSC both begin to increase, with the steeper rise in MSC reflecting the cost of congestion. With free access to the roadway, the equilibrium level of traffic will be point q_2 . Unfortunately, at this level of traffic, the MSC exceeds the marginal benefit of road use, as measured by the demand curve, by the amount [a-c]. In contrast, the socially optimal level of consumption is q_3 with price b . The congestion tax or toll is used to move the market from q_2 to q_3 by imposing a toll that is equal to the difference between the MSC and MPC that maximizes social benefit, which in this example means the toll is set equal

to [b-c].

Our problem differs from this simple scenario in three ways. First, both of the situations we investigate leave at least one lane untolled.¹ This places our problem in the class of problems economists refer to as “second-best,” in contrast to a first-best solution, which is one in which all lanes are tolled [3]. Second, our model does not allow for the possibility that use of the border crossing will decline in the face of increased user cost; rather we simply present a toll as a way to manage the current level of usage and better utilize the facilities. Third, at the border there is a much smaller range of traffic for which congestion is zero; essentially, congestion begins any time there is more than one truck in the queue, so the divergence between MPC and MSC occurs immediately. In all three of these cases, the efficiency gains from tolling are smaller than in the theoretical case presented above.

While these results provide an important theoretical foundation for understanding the rationale behind congestion pricing, current practice – as discussed in the previous section – does not necessarily set tolls at the social optimum. There are several practical difficulties with directly applying the optimal formulae in a dynamic framework. For example, the inter-arrival and service time distributions are unlikely to remain unchanged long enough to

¹This is sometimes referred to as “value pricing” and is a particularly palatable option from a political perspective since it gives users the option to avoid the toll.

by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

enable the toll to converge to its optimal value. In addition, the algorithms necessary to determine the optimal toll may require dramatic jumps in the toll charged to users, which is likely to lead to user dissatisfaction. Practical applications of dynamic pricing are therefore likely to use heuristic toll-setting algorithms which increase, but do not necessarily maximize, economic efficiency.

CURRENT BORDER OPERATIONS

While we are interested in the general applicability of congestion pricing at any border crossing, a specific example – the southbound Pacific Highway commercial border crossing – was chosen to probe the issues discussed in the previous sections. Relying on on-site observation and data collection, we constructed a simulation model of the crossing as it currently exists, and then gauged its performance under a range of congestion pricing options.

There are currently three booths for the general inspection of trucks as they arrive at PHC. One of these booths is restricted to trucks enrolled in the Free and Secure Trade, or FAST, program co-implemented by Canada, the United States, and Mexico; trucks not enrolled in FAST are restricted to the other two booths. The FAST program, which requires the carrier, driver, and importer to be registered participants of FAST, fo-

cuses on providing certification to “low risk” participants. Since the wait times at the FAST booth are significantly shorter than at the general-purpose booths, FAST participants – who have taken additional steps and have therefore incurred additional costs to enhance U.S. security – are effectively rewarded for their efforts with a shorter wait time.

Our model incorporates approximately sixteen hundred linear meters of the approach to the border. For twelve hundred meters of this approach, non-FAST trucks are restricted to the curb lane; during the remaining four-hundred-meter approach, these same trucks are routed through a series of lane splitting and re-merging before being directed into two lanes approaching the general-purpose booths. FAST traffic bypasses any queued non-FAST trucks in a passing lane that is shared with non-truck traffic; four hundred meters from the border, the remaining non-truck traffic splits off to a separate lane and FAST trucks continue on a lane dedicated for their use.

To validate and calibrate the simulation model constructed to represent the traffic flow described above, we obtained data from a comprehensive survey of inter-arrival times, service times, and queue lengths conducted over four nine-hour days in June 2006 [4]. For each hour of each of the four days for which data were available, probability distributions were fitted to the recorded

*by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.*

data; the resulting distributional parameters were then used as input to simulate each of the four days. Each day was simulated thirty times with different random number seeds, and the mean and standard deviation of the hourly queue lengths for non-FAST trucks (the only queue length recorded by the survey) were calculated. With few exceptions, the observed mean hourly queue lengths were well within the expected range of queue lengths generated by the simulation; we therefore considered the simulation model of the current border configuration to be valid.

FUTURE BORDER OPERATIONS

While there are many possible lane configurations consistent with a congestion pricing implementation at the border, we focused on two possibilities which require minimal infrastructure changes. The first scenario involves the conversion of the current FAST lane to a combined FAST/pay lane; this would leave two general-purpose “free” lanes for all other freight traffic. The second scenario would leave the FAST lane unchanged, but would convert a current general-purpose lane to a toll lane. Under both scenarios, trucks unable or unwilling to enroll in FAST would still have the option to pay for a shorter wait time.

Due to the improved responsiveness of a dynamic pricing system rela-

tive to a time-based pricing system, we only considered dynamic pricing. This would require continuously-updated information boards on the approach to the border stations listing the following information: the expected wait time for each type of lane, and the current price per vehicle for joining the toll lane. This information board should be available to trucks no later than the beginning of the sixteen-hundred-meter approach to the border. Under the combined FAST/toll scenario, no lane reconfiguration is necessary; both FAST and non-FAST toll-paying trucks would simply follow the route now taken by the FAST trucks. The information board would display two wait times – one for the two “free” lanes and one for the FAST/toll lane – and one toll price.

For the dedicated toll lane scenario, some degree of lane reconfiguration would be necessary. There are multiple ways to reconfigure the transition points for three distinct lanes; the reconfiguration considered here tries to minimize the physical changes necessary. The remaining general-purpose lane would retain its current configuration for the first twelve hundred meters, but would now feed only one lane rather than two lanes for the remaining four-hundred-meter approach to the border. Paying trucks as well as FAST trucks would follow the current FAST path for the first twelve hundred meters, and then branch off into separate lanes for the re-

by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

maintaining four hundred meters.

A method for collecting tolls would be necessary for both scenarios. While manual toll collection remains an option at many tolled transportation facilities, the preferred current collection method requires each tolled vehicle to establish an account with the tolling agency and to carry a transponder in the vehicle; when the vehicle enters the tolled facility, the transponder relays an electronic signal and the user is automatically charged. For many congestion pricing systems, such as the HOT lanes referred to earlier, this is the only payment option available. Since no physical transaction is required, service times are not increased and the process of collecting the toll does not add to the existing congestion. This method does require users to decide in advance, by registering, that they might want to use the tolled facilities, but they can postpone the final decision on a particular trip until they approach the border.

EXPERIMENTAL DESIGN

Three different border configurations were therefore examined for the simulation study: the current configuration, a configuration with two general-purpose lanes and a shared FAST/toll lane, and a configuration with one general-purpose lane and separate FAST and toll lanes. Each of the latter two scenarios used a simple toll-updating algorithm concep-

tually similar to the one originally devised for the I-15 HOT lanes. Just as the I-15 HOT lanes adjusted tolls to encourage efficient use of the HOT lanes but also to ensure that travel conditions did not deteriorate below a specified level of service, border tolls are set to minimize the probability that the waiting time for trucks in the shared or separate toll lane would exceed a specified target waiting time. Simply put, if the expected wait for an arriving truck – which is itself a factor of the current queue length, estimated service rate, and current estimated arrival rate – exceeds the target wait by a certain threshold, the toll is increased by a specified increment; if it is below the target wait by the same threshold, the toll is decreased by the same amount.

Figure 2 illustrates how the tolling algorithm works for one simulated day in the shared FAST/toll lane scenario. Both the current toll and the actual cost of waiting for each user of the shared FAST/toll lane are plotted on the graph as a function of time; the dark vertical bars each represent an instance where a non-FAST enrolled trucker paid the toll to join the shared FAST/toll lane. Since, for this example, the cost of waiting per hour was fixed to two hundred dollars per hour for all users, the waiting cost is a perfect stand-in for waiting time. Prior to 8:00 a.m. – time zero on the graph – a buildup of truckers in the FAST lane has increased the toll and effectively “priced

by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

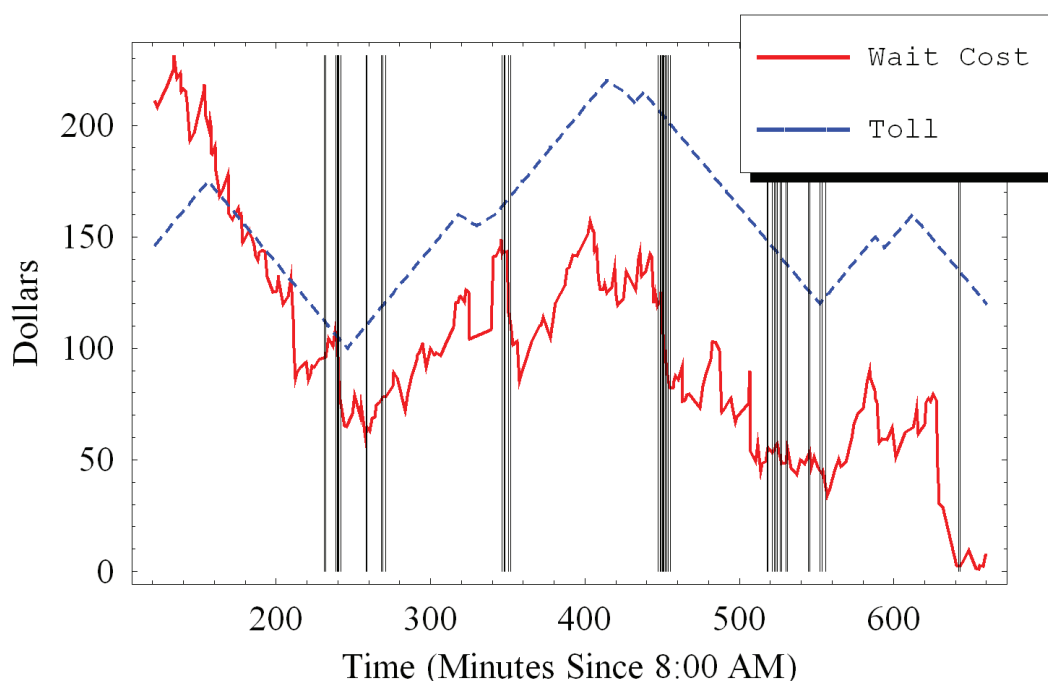
out” any non-FAST enrolled truckers. As the FAST queue declines and the wait time decreases, the toll drops as well until several non-FAST enrolled truckers opt for the FAST lane. Eventually, this leads to an increase in wait time and the toll, which once again shuts out non-FAST enrollees. Over the course of the day, the actual wait rarely exceeds the target maximum wait of sixty minutes (two hundred dollars in waiting cost), as the toll effectively regulates non-FAST arrivals to the shared FAST/toll lane.

Parameters affecting the pricing algorithm’s performance include the time interval for updating; the desired maximum target waiting time for trucks in

the shared or separate toll lane, and the desired maximum probability of exceeding that wait; the incremental amount for increasing or decreasing the toll; the threshold difference required for updating the toll; and the method used to estimate the future expected wait. Several of the parameters control the speed with which the tolls are changed; a more rapidly changing toll may therefore more accurately reflect the current state of the system, but may also be more likely to antagonize users.

To measure the performance of the current configuration and the two proposed alternatives, we focused on the average cost per user per day, where cost

Figure 2. Toll and Cost of Waiting for Shared FAST/Toll Lane Over Time



An Investigation of
Congestion Pricing
Options for
Southbound Freight
at the Pacific
Highway Crossing

by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

was defined as the cost of waiting plus the cost of any toll paid by that user. The cost of waiting was based on the total waiting time experienced by the truck and the hourly cost of time assigned to the truck at the beginning of the simulation. In the border configurations with a tolling option, trucks approaching the border are informed via an electronic signboard of the toll price and the expected waiting times in both the general-purpose lane(s) and the tolled lane. Drivers are assumed to be rational, in that they choose to pay the toll if their expected total cost after paying the toll is less than their expected cost of waiting in the general-purpose lane(s).

Each run of the simulation covered a single eleven-hour day. For each day, the period of interest was the nine-hour stretch from 8:00 a.m. to 5:00 p.m., the period for which survey data were available to calibrate the model; the two hours prior to this were used simply to initialize the queues and toll-setting algorithm, and the data collected during this initialization phase were discarded. Inter-arrival and service time distributions were based upon an “average” day as evidenced by the survey data. Environmental parameters which were varied from day to day included an “arrival factor” which, when applied as a multiplier to the mean arrival rates throughout the day, could be used to model slow or busy days; the fraction of arriv-

ing trucks enrolled in FAST; the fraction of arriving trucks which were classified as “heavy” and therefore took long to move through the radiation portal monitors; the fraction of trucks parking to conduct business prior to crossing the border; the average hourly cost of time for a truck; and the coefficient of variation (CV) of the hourly cost of time for arriving trucks.

Two separate analyses were conducted; one comparing the current configuration to the shared FAST/toll lane scenario, and the second comparing the current configuration to the separate FAST and toll lane scenario. Each analysis took place in two stages. Experiments in the first stage were used to identify those environmental parameters which impacted the relative performance of the tolling scenario under consideration; during the experimental runs, each environmental parameter was set at either a high or low level which enclosed the actual levels observed during the four-day data survey. Critical parameters identified in this stage were then included in the second experiment for each scenario; these experiments focused on the impact of the algorithmic parameters on system performance.

Table 1 shows the high and low settings used for the environmental and algorithmic parameters varied in the experiments. For average dollar value of time, the wide range chosen represents the range of estimates obtained

²Caltrans Office of Transportation Economics (http://www.dot.ca.gov/hq/tpp/offices/ote/Benefit_Cost/benefits/travel_time/index.html). Other sources give slightly lower estimates; a 2003 Oregon DOT publication suggests \$30.43/hour and a 2003 revision of a 1997 US DOT report estimates \$18.01. The higher Caltrans estimate is explained by the inclusion of an adjustment for the time cost of freight in addition to the time value of the driver.

by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

from the literature.² For most of the remaining environmental parameters the range defined by the low and high settings was meant to enclose the values observed during the four-day data survey; for the algorithmic parameters, the settings were chosen based on existing algorithms and the results of preliminary simulation runs.

Table 1. Experimental Parameters

Environmental Parameters		
Parameter	Low Setting	High Setting
Arrival Factor	0.8	1.0
% FAST	25%	35%
% Heavy	68%	88%
% Parking	5%	10%
Avg \$/Hour	\$50	\$200
CV \$/Hour	0.05	0.25
Algorithmic Parameters		
Parameter	Low Setting	High Setting
Maximum Wait (minutes)	30	60
% Less Than Maximum Wait	90%	98%
Smoothing Parameter	0.05	0.20
Update Interval (minutes)	1	6
\$ Increment	1	5
Minimum Difference (\$)	5	15

It should be noted that two key environmental parameters were set such that their high value coincided with

the observed average values: the arrival factor and the percentage of FAST enrollees. An arrival factor of one yields a daily arrival profile equal to that of an “average” day, while thirty-five percent represents the average fraction of FAST enrollees observed during the data survey. The high levels of these parameters were set to their average observed values in order to ensure the stability of all simulated days; either higher overall arrival rates or a higher overall percentage of FAST enrollees would overwhelm the FAST lane as trucks arrived faster than they could be processed. Thus, although wait times in the FAST lane are currently much shorter than in the general-purpose lanes (wait times in the FAST lane are generally a half hour or less, while those in the general-purpose lanes average well over an hour and can easily exceed two hours), the FAST lanes are currently operating close to capacity. A small increase in either parameter resulting in an increase in FAST lane arrivals would, without further mitigation, yield a dramatic increase in FAST waiting time during the peak arrival time of the day. To ensure that all simulated days were stable, therefore, the high level of each parameter was set at its current average value.

INTERPRETING THE RESULTS

The first stage of the experiment, which focused on the impact of envi-

by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

ronmental factors on the efficacy of the alternate tolling scenarios, showed that, in terms of total cost per truck, there was indeed a significant difference between the current configuration and the two tolling scenarios. Total cost per truck, which averaged \$115 across all trucks in the current configuration, dropped to \$99 in the shared FAST/toll lane configuration, but rose to \$247 in the separate FAST/toll lanes scenario. Summary results showing the total cost per truck and other daily performance measures are provided in Table 2.

Key to understanding these results is realizing that in the shared FAST/toll lane scenario, a small amount of excess capacity – as defined by the target maxi-

mum wait – is being made available to non-FAST participants. The amount of capacity being made available is not large; as mentioned earlier, the FAST lane is operating just below that level of capacity where the waiting time begins to increase dramatically. Since in this first experiment the toll is adjusted to ensure FAST participants a maximum wait of forty-five minutes, only six percent of the non-FAST participants choose to pay the toll and join the FAST/toll lane. Thus the average wait time (and waiting cost) of FAST participants goes up slightly but remains small, while the average wait time and cost of all non-FAST participants declines: those choosing to pay the toll experience significantly

Table 2. Performance of Alternative Border Scenarios

Quantity	BORDER SCENARIO		
	No Tolloed Option	Shared FAST/ Toll Lane	Separate Toll/ FAST Lanes
All Participants			
Average Waiting Cost/Truck	115.02	95.26	162.36
Average Toll/Paying Truck	0.00	65.62	270.41
Average Total Cost/Truck	115.02	99.50	246.51
FAST Participants			
Average Wait/Truck (min)	2.81	7.03	2.68
Average Waiting Cost/Truck	7.30	17.27	7.57
Non-FAST Participants			
Average Wait/Truck (min)	77.99	60.09	113.90
Average Waiting Cost/Truck	160.61	129.47	233.36
Average Total Cost/Truck	160.61	135.08	355.96
Average Total Toll Revenue	0.00	3,428.45	45,902.60
Average % Choosing Toll	0.00%	5.97%	44.22%

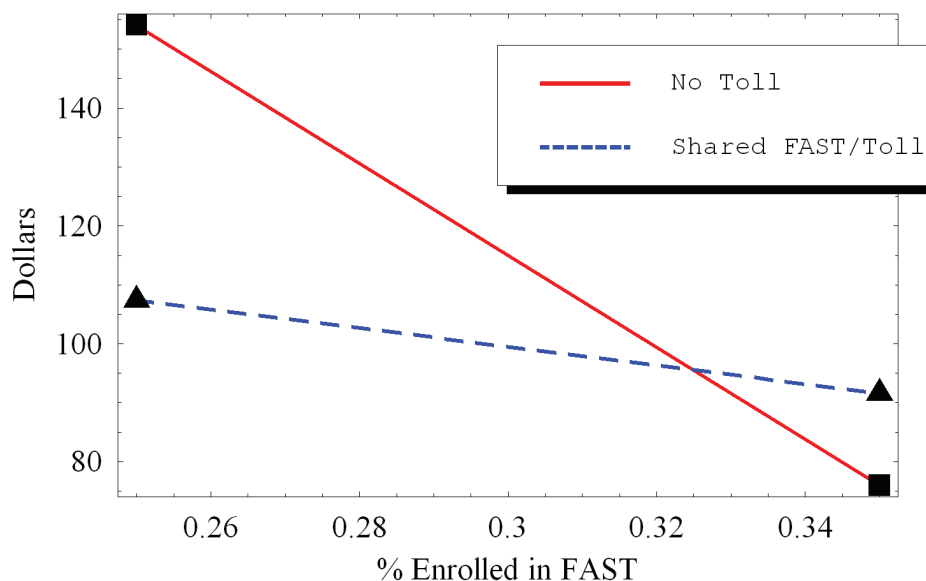
by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

shorter wait times, while those choosing not to pay have fewer trucks in front of them and also see their wait time and cost decrease. Since the decline in non-FAST waiting cost is more than the toll paid averaged over all FAST participants, the total cost per non-FAST participant declines by nearly twenty-five dollars, which more than cancels out the ten dollar increase in cost experienced by the smaller number of FAST users.

In moving from the current configuration to the separate FAST/toll lanes scenario, on the other hand, capacity is effectively removed from the border crossing since the toll is regulated to ensure a maximum waiting time; when the likely waiting time exceeds the target amount, the toll increases until ar-

rivals are discouraged from joining the tolled lane. Thus only slightly more than forty-four percent of non-FAST participants choose to join the tolled lane, rather than the fifty percent expected to join each non-FAST lane if no toll were charged. While it may seem that a 44/56 split of all non-FAST participants between two lanes would not yield much of an average waiting time difference from the 50/50 split of the same trucks in the non-tolled scenario, elementary queueing theory does indeed show that such an uneven split will increase average waiting time across both lanes, especially as the system approaches capacity: the decrease in waiting time experienced by those in the tolled lane will be more than offset

Figure 3. Interaction Between Tolling Option and FAST Enrollment



by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

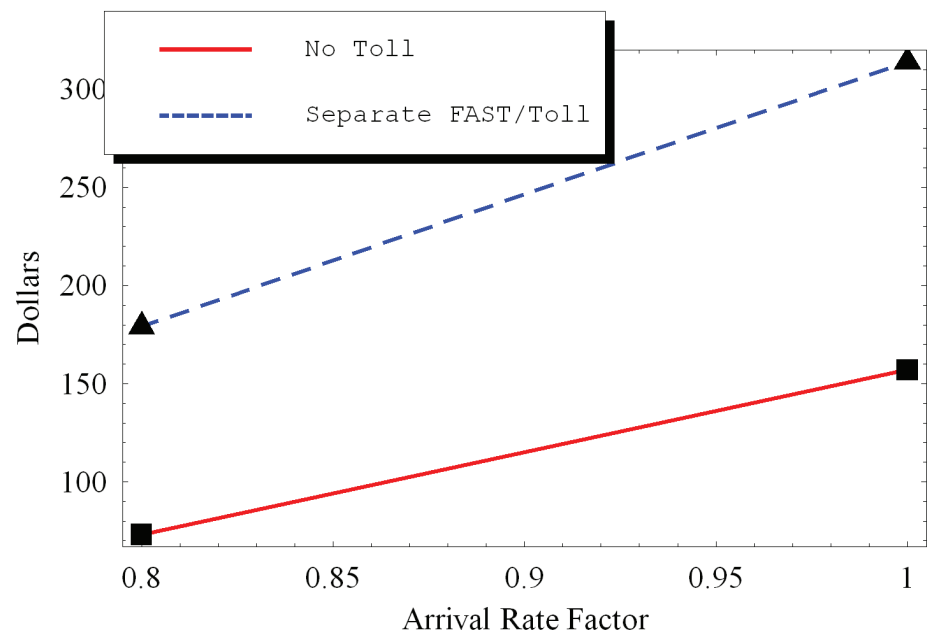
by much greater increases for those in the remaining general-purpose lane. Compounding the difficulty for non-FAST enrollees, the half that choose to pay the toll incur an even higher total cost. Since there is so much additional waiting for non-FAST enrolled trucks, especially in the general-purpose lane, the toll – which is meant to capture the cost of waiting imposed on others – is much higher than in the shared FAST/toll lane scenario.

Furthermore, analysis of the interactions between the environmental parameters and the choice of the tolling scenario revealed a significant interaction between the use of each tolling scenario and certain environmental parameters. For the shared FAST/toll lane

scenario, total cost per truck decreased by nearly fifty dollars in switching from the current configuration to the tolled scenario when FAST enrollment was low, but increased by over fifteen dollars when enrollment was high (Figure 3). Thus the gains noted above for the shared FAST/toll scenario over the current configuration are only evident when FAST enrollment is low and there is spare capacity in the FAST lane; at the current enrollment rate for FAST at PHC, a lack of excess capacity leads to a slight increase in total cost as non-FAST participants pay to join the FAST lane.

For the separate FAST/toll lanes scenario, the overall arrival rate significantly impacts the relative performance of the tolled option; while the average

Figure 4. Interaction Between Tolling Option and Arrival Rate



by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.

cost per truck generally increases in moving from the current configuration to the tolled scenario, that increase is fifty percent greater at current arrival rates than at a setting equal to eighty percent of the current values (Figure 4). As with the shared FAST/toll option, the separate FAST/toll lane option appears more palatable in a less-congested system than currently exists at the PHC.

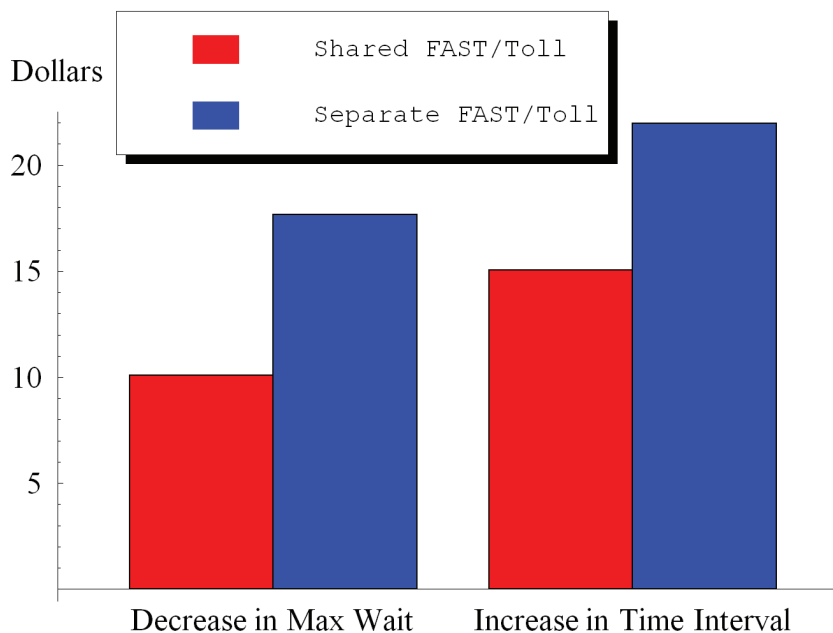
The second stage of the analysis, which looked at the impact of algorithmic parameters on the performance of each tolling scenario, yielded, for the most part, expected results (Figure 5). For both tolled scenarios, the target maximum waiting time was a significant factor; lowering the target maximum wait from sixty to thirty minutes

priced out many potential toll-paying users and raised the average cost per truck for both tolled scenarios. Increasing the time interval for updating the toll from one to six minutes, also, on average, significantly increased the average total cost per truck. As expected, frequent updates allow a closer match between the toll being charged and the current border status, thus helping to maximize the effectiveness of both tolling systems.

CONCLUSIONS AND POLICY RECOMMENDATIONS

The primary conclusion from our analysis is the importance of capacity in any politically acceptable congestion

Figure 5. Impact of Algorithmic Parameters on Increasing Average Cost



An Investigation of
Congestion Pricing
Options for
Southbound Freight
at the Pacific
Highway Crossing

*by Matthew R. Roelofs, Ph.D.
and Mark Springer, Ph.D.*

pricing implementation. Where excess capacity in the FAST lane is converted to usable space via a toll, performance measures for the system improve. Current usage of the FAST lane, however, leaves little excess capacity for toll-paying non-FAST participants, and given that FAST enrollment is encouraged by the CBP and may be expected to increase, implementing a toll via a shared FAST/toll lane is unlikely to offer much advantage for toll-paying trucks or to lower the average cost experienced by all trucks. Furthermore, in the case where the limited excess FAST lane capacity is not tapped and one of the two current general-purpose lanes is converted into a toll lane, performance of the system—as measured by the waiting and tolling cost per truck—degrades across all likely parameter combinations.

Two possible carrier reactions to a toll, neither of which is considered in this analysis, could improve a toll's relative performance. The current analysis assumes that the number of trucks crossing the border and their arrival pattern remains unchanged regardless of the tolling scenario; it also assumes that tolling does not encourage users to either abandon or adopt FAST enrollment. Analyzing the impact of relaxing either or both of these assumptions, which would require further detailed knowledge of the carriers' demand functions, could increase the relative performance of either or both

tolling scenarios.

In general, however, the results reported here mimic current experience with congestion pricing plans and implementations across the country. The plans most likely to successfully move beyond the planning stage have been those that have made additional capacity available for a fee; both SR91, where new express lanes were tolled from the beginning of their construction, and I-15, where excess capacity was opened to paying users on an underutilized HOV lane, took this approach. Attempts to charge for currently "free" transportation resources without a corresponding increase in capacity have rarely progressed beyond the planning stage in the United States. Given the absence of much excess capacity in the FAST lane at PHC, a tolled lane does not seem to be a viable option if a key performance measure is the average cost born by all users. A possible third alternative, not considered here because of the infrastructure requirements, would be to open a fourth booth exclusively for tolling. Indeed, given the current high utilization and likely continued increase in cross-border traffic, some further increase in PHC capacity is likely to be warranted in the near future; adding a tolled option in the form of a fourth booth should certainly be one of the scenarios examined when additional capacity is considered.

REFERENCES

- [1] USDOT, *Washington State-British Columbia IMTC ITS-CVO Border Crossing Deployment Evaluation Final Report*, October 2003.
- [2] Whatcom Council of Governments, *Pacific Crossing Commercial Vehicle Border Delay Survey*, November 2006.
- [3] Verhoef, Eric, Peter Nijkamp, and Piet Rietveld, "Second-Best Congestion Pricing: The Case of an Untolled Alternative" *Journal of Urban Economics* (vol. 40, pp. 279-302, 1996).
- [4] Whatcom Council of Governments, *Pacific Crossing Commercial Vehicle Border Delay Survey*, November 2006.

Research Report No. 4
August 2007

Border Policy Research Institute
Western Washington University
Bellingham, Washington
www.ac.wvu.edu/~bpri/

