

# The Booth Tolls for Thee

Adam Chandler

Matthew Mian

Pradeep Baliga

Duke University

Durham, NC

Advisor: William G. Mitchener

## Summary

We determine the optimal number of tollbooths for a given number of incoming highway lanes. We interpret optimality as minimizing “total cost to the system,” the time that the public wastes while waiting to be processed plus the operating cost of the tollbooths.

We develop a microscopic simulation of line-formation in front of the toll-booths. We fit a Fourier series to hourly demand data from a major New Jersey parkway. Using threshold analysis, we set upper bounds on the number of tollbooths. This simulation does not take bottlenecking into account, but it does inform a more general macroscopic framework for toll plaza design.

Finally, we formulate a model for traffic flow through a plaza using cellular automata. Our results are summarized in the formula for the optimal number  $B$  of tollbooths for  $L$  lanes:  $B = \lfloor 1.65L + 0.9 \rfloor$ .

## Previous Work in Traffic Theory

Most models for traffic flow fall into one of two categories: microscopic and macroscopic.

Microscopic models examine the actions and decisions made by individual cars and drivers. Often these models are called *car-following models*, since they use the spacing and speeds of cars to characterize the overall flow of traffic.

Macroscopic models view traffic flow in analogy to hydrodynamics and the flow of fluid streams. Such models assess “average” behavior, and commonly-used variables include steady-state speed, flux of cars per time, and density of traffic flow.

---

*The UMAP Journal* 26 (3) (2005) 283–297. ©Copyright 2005 by COMAP, Inc. All rights reserved. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice. Abstracting with credit is permitted, but copyrights for components of this work owned by others than COMAP must be honored. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior permission from COMAP.

Some models bridge the gap between the two kinds, including the gas-kinetic model, which allows for individual driving behaviors to enter into a macroscopic view of traffic, much as ideal gas theory can examine individual particles and collective gas [Tampere et al. 2003].

The tollbooth problem involves no steady speed, so macroscopic views may be tricky. On the other hand, bottlenecking is complex and tests microscopic ideas.

An  $M|M|n$  queue seems appropriate at first: Vehicles arrive with gaps (determined by an exponential random variable) at  $n$  tollbooths, with service at each tollbooth taking an exponential random variable amount of time [Gelenbe 1987]. We incorporate aspects of the situation from a small scale into a larger-scale framework.

## Properties of a Successful Model

A successful toll-plaza configuration should

- maximize efficiency by reducing customer waiting time;
- suggest a reasonably implementable policy to toll plaza operators; and
- be robust enough to handle efficiently the demands of a wide range of operating capacities.

## General Assumptions and Definitions

### Assumptions

- All drivers act according to the same set of rules. Although the individual decisions of any given driver are probabilistic, the associated probabilities are the same for all drivers.
- Bottlenecking downstream of the tollbooths does not hinder their operation. Vehicles that have already passed through a tollbooth may experience a slowing down due to the merging of traffic, but this effect is not extreme enough to block the tollbooth exits.
- The number of highway lanes does not exceed the number of tollbooths.
- All tollbooths offer the same service, and vehicles do not distinguish among them.
- The amount of traffic on the highway is dictated by the number of lanes on the highway and not by the number of tollbooths. Changing the number of tollbooths does not affect “demand” for the roadway.
- The number of operating tollbooths remains constant throughout the day.

## Terms and Definitions

- A “highway lane” is a lane of roadway in the original highway before and after the toll plaza.
- *Influx* is the rate (in cars/min) of cars entering all booths of the plaza.
- *Outflux* is the rate (in cars/min) of cars exiting all booths of the plaza.

## Optimization

We seek to balance the cost of customer waiting time with toll plaza operating costs.

- The daily cost  $C$  of a tollbooth is the total time value of the delays incurred for all individuals (driver plus any passengers) plus the cost of operation of the booth. The tolls and the startup cost of building the plaza are not part of this cost.
- $a$  is the average time-value of a minute for a car occupant.
- $\gamma$  is the average car occupancy.
- $N$  is the total number of (indistinct) tolls paid over the course of one day.
- $L$  is the number of lanes entering and leaving a plaza.
- $W$  is the average waiting time at a tollbooth, in minutes.
- $B$  is the number of booths in the plaza.
- $Q$  is the average daily operating cost of a human-staffed tollbooth.

We seek a number of toolbooths  $B$  that minimizes cost  $C$ .

The total waiting time per car is  $WN$ , so the total cost incurred by waiting time is  $WN\gamma$ . General human time-value is cited as \$6/hour or  $a = \$0.10/\text{min}$  [Boronico 1998]. The cost to operate a booth for a day is  $QB$ . The average annual operation cost for a human-staffed tollbooth is \$180,000, so we set  $Q = \$180000/365.25 \text{ days}$  [Sullivan et al. 1994].

Reasoning that  $W$  depends on  $B$ , we have

$$C(B) = WN\gamma + QB.$$

## Car Entry Rate

We fit a curve to mean hourly traffic-flow data from Boronico [1998]. To interpolate an influx rate for every minute during the day, we fit a Fourier series approximation, whose advantage is its periodicity, with period of one day. [EDITOR'S NOTE: We omit the table of data from Boronico [1998] and the 17-term expression for the approximating series.]

# Model 1: Car-Tracking Without Bottlenecks

## Approach

We seek an upper bound on the optimal number of booths for a particular number of lanes.

## Assumptions

- Each vehicle is looking to get through the toll plaza as quickly as possible, and the only factor that may cause Car A, which arrives earlier than Car B, to leave later than B is the random variable of service time at a tollbooth. In other words, cars do not make bad decisions about their wait times
- Customers are served at a tollbooth at a rate defined by an exponential random variable (a common assumption in queueing theory [Gelenbe 1987]) with mean 12 s/vehicle, or 5 cars/min.
- Traffic influx occurs on a “per lane” basis, meaning that influx per lane is constant over all configurations with varying number of lanes.
- Bottlenecking occurs more frequently when there are more tollbooths, given a particular number of lanes. This implies that omitting bottlenecking from our model will cause us to overestimate the optimal number of tollbooths.
- There exists a time-saving threshold such that if the waiting time saved by adding another tollbooth is under this threshold, it is not worth the trouble and expense to add the tollbooth. We assume that if an additional tollbooth does not reduce the maximum waiting time over all cars by the same amount as the average time that it takes to serve a car at a tollbooth (12 s = 0.2 m), then it is unnecessary.
- An incoming car chooses the tollbooth that will be soonest vacated, if all are currently occupied. If only one is vacant, the car chooses that tollbooth. If multiple tollbooths are vacant, the car chooses the one vacated the earliest.
- Cars make rational decisions with the goal of minimizing their wait times.

## Expectations of the Model

- An additional booth should not increase waiting time.
- Each additional tollbooth offers diminishing returns of time saved.

## Development of Model

Cars arrive at the toll plaza at a rate described by the Fourier series approximation of the data from Boronico [1998]. Cars are considered inside the toll plaza (meaning that we begin to tabulate their waiting times) when they are either being served or waiting to be served.

Service time does not count as waiting time; so if a car enters the toll plaza and there is a vacant tollbooth, its waiting time is 0. If there are no vacant tollbooths, cars form a queue to wait for tollbooths, and they enter new vacancies in the order in which they entered the toll plaza. Once a car has been served, it is considered to have exited its tollbooth and the toll plaza as a whole.

Our car-tracking model does not factor in the cost of tollbooth operation.

## Simulation and Results

For  $L$  highway lanes,  $L \in \{1..8, 16\}$ , we ran the simulation for numbers  $B$  of tollbooths up to a point where additional booths no longer have any noticeable effect on waiting time. We exhibit results for a 6-lane highway in **Table 1**.

**Table 1.**  
Waiting times (in minutes) for six-lane simulations.

Booths	Ave. wait	Ave. wait for wait > 0	Max. wait	Marginal utility
6	28	43	99	N/A
7	12	28	55	44
8	6	17	32	23
9	2	8	16	16
10	0.25	1.22	2.78	13
11	0.02	0.17	0.75	2
12	0.004	0.07	0.31	0.44
13	0.001	0.04	0.27	0.04

The column “Marginal Utility” shows how much each additional booth reduces maximum waiting time. For the 13th booth, this value is 0.04 min. To choose an optimal number of booths by threshold analysis, we seek the first additional booth that fails to reduce the maximum waiting time for a car by at least the length of the average tollbooth service time (0.2 min). So, based on our assumptions, it is unnecessary to build a 13th tollbooth for a toll plaza serving 6 lanes of traffic. Thus, we set  $B = 12$  for  $L = 6$ . **Table 2** shows the optimal number of tollbooths for other various numbers of lanes.

**Table 2.**  
Optimized number of booths.

Lanes	1	2	3	4	5	6	7	8	16
Booths	4	5	7	8	10	12	13	16	29

We also explore the situation of one booth per lane. Regardless of the number of lanes, we find average wait times of around 30 min (over 40 min for cars that wait at all), and maximum wait time of around 100 min.

## Discussion

Our results match our expectations. The optimal number of booths increases with the number of lanes, each additional booth reduces waiting time, and additional booths yield diminishing returns in reducing waiting time.

The benefits of this rather simplistic model are its speed and the definite upper bounds that it offers.

## Model 2: Cost Minimization

### Approach

This model is concerned less with the details of individual vehicular motion and decision-making than with the general aggregate effect of the motions of the cars. We monitor traffic over the course of one day.

For instance, there is no need for this model to decompose analytically the situation of two cars trying to merge into the same lane. Instead, it recognizes that beyond a certain threshold of outflux from the booths, some bottlenecking will occur.

Also, this model addresses the cost of daily operation of the plazas.

### Assumptions, Variables, and Terms

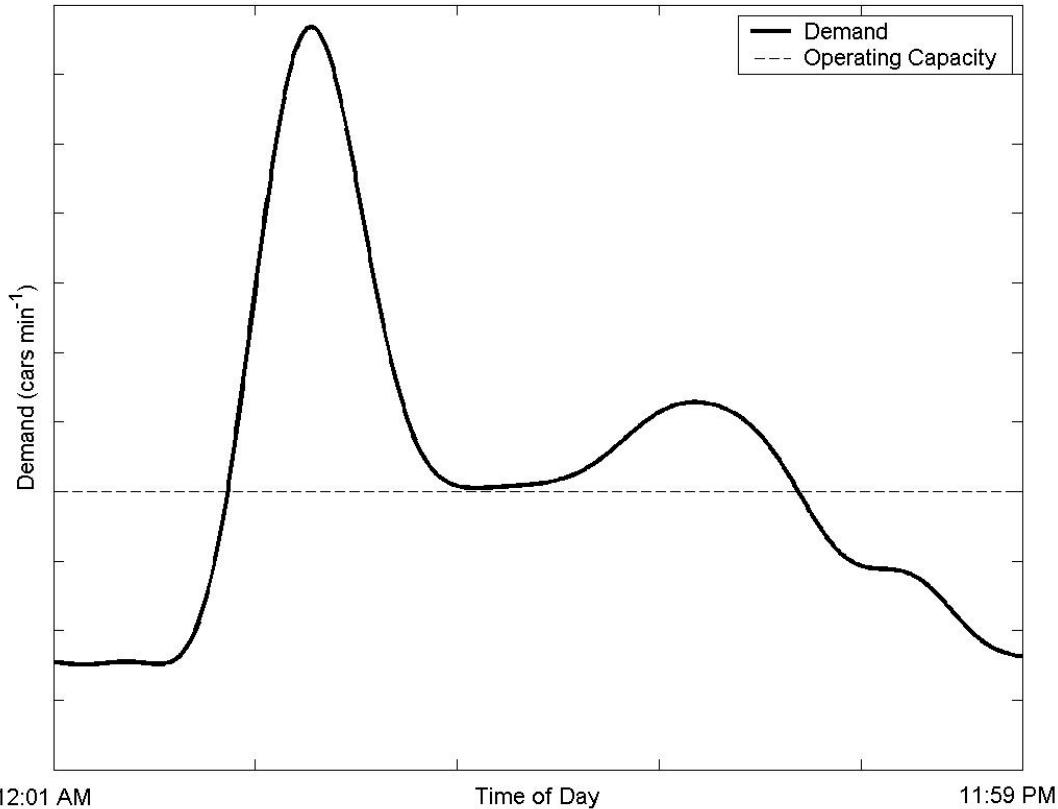
- The average waiting time per car in the toll plaza,  $W$ , is comprised of time in line ( $W_1$ ), service time ( $W_2$ ), and bottlenecking ( $W_3$ ).
- $F_i$  (a function of time) is the influx (cars/min) to the plaza from one lane,  $F_o$  is the outflux (cars/min) from the booth.
- $r$  is the maximum potential service rate (cars/min).
- There is an outflux barrier,  $K$  (cars/min), above which bottlenecking takes place. We take it to be linear in  $L$  and independent of  $B$ , and we call it the *bottlenecking threshold*.

## Development

From the definitions, we have  $W_2 = 1/r$ . Both  $W_1$  and  $W_3$  depend on  $B$ .

The average time in line,  $W_1$ , begins to accumulate when the influx of traffic exceeds the toll plaza capacity (see **Figure 1**). The influx is  $LF_i(t)$  and the

maximal rate of service is  $B_r$ . We integrate over time to calculate how many cars were forced to wait in line.



**Figure 1.** The area below the curve and above the line represents cars in line.

Integrating again (over time) gives us the total waiting time for all those cars (with 3600 as a scale factor for time units), and dividing by the total number of cars gives the average waiting time:

$$W_1 = \frac{3600}{N} \int_0^{24} \int_0^t \max(LF_i(\tau) - Br, 0) d\tau dt.$$

We obtain  $W_3$  in similar fashion:

$$W_3 = \frac{3600}{N} \int_0^{24} \int_0^t \max(F_o(\tau, B) - K, 0) d\tau dt.$$

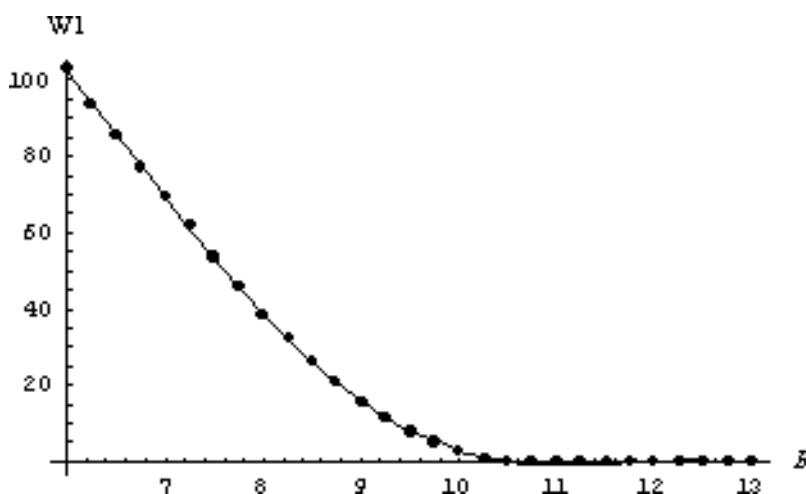
The problem is to determine the variable(s) that  $K$  depends on. First,  $K$  is not directly dependent on  $B$ , since bottlenecking should only be a result of general outflux from the booths into  $L$  lanes. Instead,  $K$  depends indirectly on the number of booths, because  $K$  depends on outflux  $F_o$ , which in turn depends on  $B$ . Also,  $K$  also can be considered a linear function of  $L$ , because  $L$  is directly proportional to influx, which, by the law of conservation of traffic, must equal outflux in the aggregate.

## Simulation and Results

We use the same data and Fourier series for traffic influx as in Model 1. We focus our attention on the case  $L = 6$ ; other values are analogous.

We use Mathematica to integrate numerically the expression for  $W_1$  for a given  $L$  and  $r = 5$  cars/min. ( $N$  comes from integration of the influx expression.) We do the calculation for values of  $B$  ranging from  $L$  to  $L + 7$  (since  $L + 7$  is usually greater than the upper bound from Model 1) with step size 0.25. We fit a quartic polynomial fit to the resulting points  $(B, W_1(B))$  to get  $W_1$  as a function of  $B$ .

We illustrate for  $L = 6$ . We find  $N = 92,355$ . We plot  $W_1$  for values of  $B$  from 6 to 13, in steps of 0.25, together with the best-fit quartic, in **Figure 2**.



**Figure 2.**  $W_1(B)$  for  $B$  ranging from 6 to 13 (actual points with quartic fit).

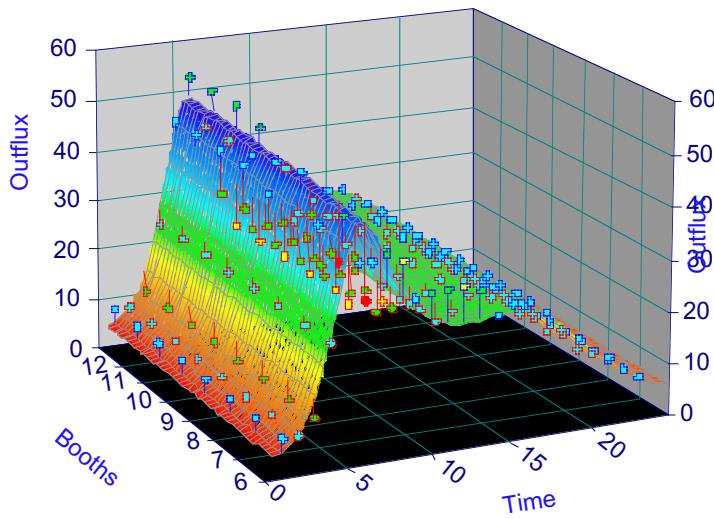
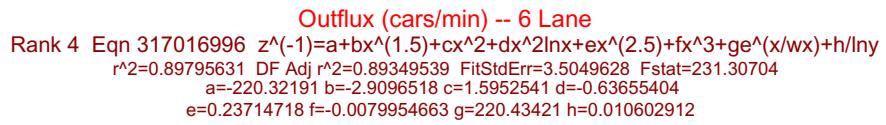
The recipe for  $W_3(B)$  is somewhat less straightforward, since  $F_o(t)$  is generated from a stochastic distribution, unlike the deterministic  $F_i(t)$ . Also,  $F_o(t)$  depends on  $B$ , a significant complication. We ran at least 20 trials of each case  $(L, B)$  under the first model; the averaged outcomes of their outflux functions are the function that we use for outflux in this model's simulation, henceforth referred to just as  $F_o(t, B)$ .

We use surface-fitting software (Systat's TableCurve3D) to generate an expression for outflux as a function of time and the number of booths and use this expression in the compound integral for  $W_3$  in integrating numerically. As before, we generate a scatter plot of points  $(B, W_3(B))$  and fit a quartic polynomial.

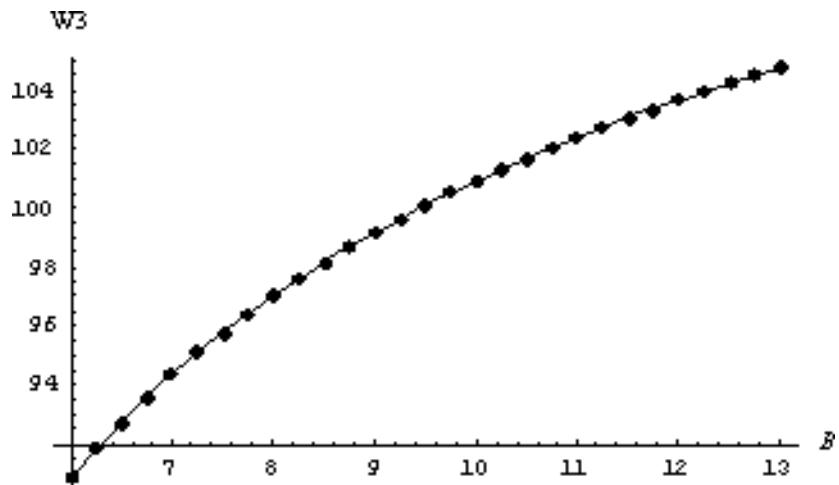
The values of  $R^2$  for the surface fits all fall between .84 and .95, which are acceptable values. All of the quartic fits have  $R^2$  near 1.

For the case  $L = 6$ , **Figure 3** shows the surface fit and **Figure 4** shows the function  $W_3$ .

With  $W_3$  a quartic polynomial in  $B$ , minimization via calculus yields a solution. For 6 lanes, Mathematica's numeric solver gives the minimum at  $B = 10.84$ . Values for various numbers of lanes are summarized in **Table 3**.



**Figure 3.**  $L = 6$ : Surface fit for outflux function in terms of time (h) and number of booths.



**Figure 4.**  $L = 6$ : Plot of  $W_3(B)$  for  $B$  ranging from 6 to 13 (actual points plus quartic fit).

**Table 3.**  
Optimized number of booths—final recommendations from Model 2.

Lanes	1	2	3	4	5	6	7	8	16
Booths	3	5	6	7	9	11	12	14	27

## Discussion

This model calculates total waiting time for drivers based on general ideas of traffic flow. The results are reasonable and satisfy many of our expectations for a successful model. The recommended  $B$ -values increase monotonically with  $L$  and are all less than the upper bounds produced in Model 1. One booth per lane is nowhere near optimal, because (as we can see from the graphs of  $W_1$  and  $W_3$ ), while bottlenecking is zero, waiting time in line is much higher, thus diminishing the effect of bottlenecking.

Given the model's success, it may be disheartening to acknowledge its lack of robustness. Any adjustments to fine-scale aspects of traffic, such as the addition of a potential E-ZPass lane (to be discussed later), would be impossible to implement. Perhaps the rate of service  $r$  could be adjusted higher for such a scenario, but changing lanes before the tollbooths would be difficult to capture with this model.

## Model 3: Cellular Automata

### Motivation

What effect does the discreteness of traffic have on the nature and solution of the problem? A continuous model of traffic may neglect the very factors that give rise to traffic congestion and jamming. To address this possibility, we turn to cellular automata theory to develop a discrete, microscopic model.

### Approach

Each cell is designated as a vehicle, a vacancy, or a barrier to traffic flow. The model follows individual vehicles through the plaza and computes the waiting time for each. The total waiting time measures the plaza's efficiency.

In any particular time step, a vehicle advances, changes lanes, or sits still. Vehicles enter the plaza from a stretch of road containing a specific number of lanes. As a vehicle approaches the string of tollbooths, the road widens to accommodate the booths (given that there are more tollbooths than lanes). There is a specific delay associated with using a tollbooth. Once a vehicle leaves a booth, it merges into a roadway with the original number of lanes.

### Assumptions

- The plaza consists of occupied cells, vacant cells, and “forbidden” cells.
- Cells represent a physical space that accommodates a standard vehicle with buffer regions on both sides.
- All vehicles are the same size.

## Governing Dynamics

Cars move through the toll plaza according to rules. Each vehicle has options, each with an associated probability. For each time step, the following rules are applied in sequential order:

1. Starting at the front of the traffic and moving backward (with respect to the flow), vehicles advance to the cell directly in front of them with probability  $p$ ; if the next cell is not vacant, the vehicle does not advance and is flagged. This probability is meant to simulate the stop-and-go nature of slowly moving traffic. We can think of  $p$  as a measure of driver attentiveness;  $p = 1$  corresponds to the case where drivers are perfectly attentive and move forward at every opportunity, while  $p = 0$  represents the extreme case where drivers have fallen asleep and fail to move forward at all.
2. Using an influx distribution function, the appropriate number of new vehicles is randomly assigned to lanes at the initial boundary (see next section).
3. Starting at the front of traffic and moving backward, vehicles flagged in step 1 are given the opportunity to switch lanes. For each row of traffic, the priority order for switching is determined by a random permutation of lanes. Switching is attempted with probability  $q$ . If switching is attempted, left and right merges are given equal probability to be attempted first. If a merge in one direction (i.e. left or right) is impossible (meaning that the adjacent cell is not vacant), then the other direction is attempted. If both adjacent cells are unavailable, the vehicle is not moved.
4. Total waiting time for the current time step is computed by determining the number of cells in the system containing a vehicle.
5. The number of vehicles advancing through the far boundary (end of the simulation space) are tabulated and added to the total output. This number is later used to confirm conservation of traffic.

## Population Considerations

The Fourier series for daily influx distribution of cars is still valid for the automata model, but the influx values must be scaled to reflect the effective influx over a much smaller time interval (a single time step). The modified influx function,  $F_{\text{in}}$ , is computed as follows:

$$F_{\text{in}}(\tau) = \min \left( \left\lfloor \frac{F_{\text{in}}(t)}{\eta} \right\rfloor, L \right),$$

where  $\eta$  is a constant factor required for the conversion from units of  $t$  to those of  $\tau$  and  $L$  is the number of initial travel lanes.

## Computing Wait Time

Wait time is determined by looking through the entire matrix at each time step and noting the number of cells with positive values. The only cells containing positive values are those representing vehicles. Thus, by counting the number of vehicles in the plaza at any given time, we are also counting the amount of time spent by vehicles in the plaza (in units of time steps).

At time step  $i$ , total cumulative waiting time is computed as follows:

$$W_i = W_{i-1} + 1(\text{plaza}(x, y) > 0),$$

where  $1()$  denotes an indicator function and  $\text{plaza}$  denotes the matrix of cells.

## Simulation and Results

The cost optimization method defines total cost as

$$C_{\text{total}} = \alpha\gamma NW(B, L) + BQ.$$

Using the cellular automata model, we compute waiting time as a function of both the number of lanes and the number of tollbooths. For fixed  $L$ , we compare all values of  $C_{\text{total}}$  and choose the lowest one. The results are presented in **Table 4**.

**Table 4.**  
Optimization for cellular automata model.

Highway lanes	Optimal number of booths	
	Typical day	Rush hour
1	1	2
2	4	4
3	5	6
4	7	7
5	8	9
6	10	11
7	12	13
8	14	15
16	27	29

Each value in **Table 4** represents approximately 20 trials. Through these trials, we noted a remarkable stability in our model. Despite the stochastic nature of our algorithm, each number of lanes was almost always optimized to the same number of tollbooths. There were a handful of exceptions; they occurred exclusively for small numbers of highway lanes (< 3 lanes).

## Sensitivity Analysis

Our cellular automata model is relatively insensitive to both  $p$  and  $q$ . Changes of  $\pm 11\%$  in  $p$  and  $\pm 5.2\%$  in  $q$  have no effect on the optimal number of tollbooths

for a six-lane highway. On the other hand, increasing the delay time by 25% shifts the optimal number of booths from 10 to 11 (10%). Decreasing the delay by 25% has no effect on the solution. Perhaps additional work could lead to an elucidation of the relation between delay and optimal booth number that could help stabilize the cellular automata model.

## Comparison of Results from the Models

**Table 5** show the optimal number of booths.

**Table 5.**  
Comparison of final recommendations for three models.

Lanes	Model		
	Car-tracking	Macroscopic	Automata
1	4	3	2
2	5	5	4
3	7	6	5
4	8	7	7
5	10	9	8
6	12	11	10
7	13	12	12
8	16	14	14
16	29	27	27

The car-tracking model serves as an upper bound for the optimal number of booths, due to its omission of bottlenecking, a fact confirmed in the table. The cellular automata model, on the other hand, incorporates bottlenecking. Due to its examination of each car and each period waited, we lean more toward the cellular automata model for a determination of the optimal number of booths that is more accurate than those of the other two models.

The optimal values for each model are fit very well ( $r^2 > .996$ ) by a straight line, with slopes between 1.6 and 1.7.

## Conclusion

We use three models—the car-tracking model, the macroscopic model for total cost minimization, and the cellular automata model—to determine the optimal (per our definition) number  $B$  of tollbooths for a toll plaza of  $L$  lanes.

The car-tracking model uses a simple orderly lineup of cars approaching tollbooths and ignores bottlenecking after the tollbooths; it provides a strong upper bound on  $B$  for any given  $L$ .

The macroscopic model looks at the motion of traffic as a whole. It tabulates waiting time in line before the tollbooths by considering times when traffic influx into the toll plaza is greater than tollbooth service time. It also

finds bottlenecking time by assuming there exists a threshold of outflux, above which bottlenecks will occur, and notices when outflux is greater than said threshold. This is a much more accurate model than the Car-Tracking Model, and it provides us with reasonable solutions for  $B$  in terms of  $L$ .

The cellular automata model looks at individual vehicles and their “per lane length” motion on a toll plaza made up of cells. With a probabilistic model of how drivers advance and change lanes, this model details far better than the other models the waiting time in line and the bottlenecking after the tollbooths.

Thus, we recommend values closer to those provided by the automata model than the macroscopic one. In order to write  $B$  explicitly in terms of  $L$ , we invoke the linearity of the results. Also, to preserve integral values for  $B$ , we use the floor function and determine that

$$B = \lfloor 1.65L + 0.9 \rfloor.$$

## Potential Extension and Further Consideration

Our models assume that all booths are identical. However, systems such as E-ZPass allow a driver to pay a toll electronically from an in-car device without stopping at a tollbooth. If all E-ZPass booths also double as regular teller-operated booths, much of our models remain the same, except that the average service rate might increase. The trouble comes when all the booths are not the same and drivers may need to change lanes upon entering the plaza. This directed lane changing was not implemented in any of the models presented here, but could easily become a part of the automata model. Exclusive E-ZPass booths also would drastically reduce the operating cost for the booth, since an operator’s salary would not need to be paid (from \$16,000 to \$180,000 annually) [Sullivan 1994].

## References

- Boronico, Jess S., and Philip H. Siegel. Capacity planning for toll roadways incorporating consumer wait time costs. *Transportation Research A* (May 1998) 32 (4): 297–310.
- Daganzo, C.F., et al. 1997. Causes and effects of phase transitions in highway traffic. ITS Research Report UCB-ITS-RR-97-8 (December 1997).
- Gartner, Nathan, Carroll J. Messer, and Ajay K. Rathi. 1992. *Traffic Flow Theory: A State of the Art Report*. Revised monograph. Special Report 165. Oak Ridge, TN: Oak Ridge National Laboratory. <http://www.tfhrc.gov/its/tft/tft.htm>.
- Gelenbe, E., and G. Pujolle. 1987. *Introduction to Queueing Networks*. New York: John Wiley & Sons.

- Jost, Dominic, and Kai Nagel. 2003. Traffic jam dynamics in traffic flow models. Swiss Transport Research Conference. <http://www.strc.ch/Paper/jost.pdf>. Accessed 4 February 2005.
- Kuhne, Reinhart, and Panos Michalopoulos. 1992. Continuum flow models. Chapter 5 in Gartner et al. [1992].
- Schadschneider, A., and M. Schreckenberg. Cellular automaton models and traffic flow. *Journal of Physics A: Mathematical and General* (1993) 26: L679–L683.
- Sullivan, R. Lee. Fast lane. *Forbes* (4 July 1994) 154: 112–115.
- Tampere, Chris, Serge P. Hoogendoorn, and Bart van Arem. Capacity funnel explained using the human-kinetic traffic flow model. <http://www.kuleuven.ac.be/traffic/dwn/res2.pdf>. Accessed 4 February 2005.



Matthew Mian, William G. Mitchener (advisor), Pradeep Baliga, and Adam Chandler.