Modeling Toll Plaza Behavior Using Queuing Theory

February 7, 2005

Abstract

When a toll plaza is designed, choosing the right number of tollbooths is a critical issue. In this paper, we try to determine the optimal number of tollbooths by creating a model for traffic in a toll plaza. After discussing the natural behavior of traffic and making a few reasonable assumptions to simplify traffic streams in a toll plaza, we break the travel process in a toll plaza into two stages: toll collection and merging. We apply Queuing Theory to each stage, modeling each stage as a queuing system. Having determined that an optimal toll plaza minimizes travel time, we derive a formula to calculate the average wasted time per driver in terms of number of incoming lanes, traffic flow, and number of tollbooths. The average wasted time is the portion of the travel time that is dependent on the parameters of the system. Using this formula we compute the average wasted time for a broad range of lane counts, traffic levels, and tollbooth numbers, from which optimal tollbooth counts for each configuration were found and are displayed in tabular form. We recommend that the number of tollbooths should be around 7 to 10 depending on the size of highway when toll collection service rate is low. We consider the scenario where there are equal numbers of tollbooths and incoming lanes, and conclude that it is optimal only when toll collection is very fast, and it is not effective in other cases.

Contents

1 Introduction

Toll financing has been used throughout the history of civilization to make the building of long-distance roads possible. Toll roads in Asia were known to ancient Greek writers in antiquity. The first turnpikes in America were built in the 1790s and helped to open the Midwest to settlement. Beginning in the 1940s, America's first modern freeways were financed with tolls, paving the way to the Interstate system Americans now enjoy. Today developing nations such as China are building their own networks of superhighways, and they too are turning to the tollbooth to foot the bill. As the congestion and pollution from too many cars on the streets become an increasing concern in many cities, another benefit of tolling has revealed itself: Tolls are being used successfully in places such as Singapore and London not just to finance road construction, but to limit the flow of vehicles into the urban core, increasing transit usage and unclogging the crowded streets [12].

Despite its many advantages, there is one undeniable drawback to tolling that is the bane of drivers and road builders alike: When traffic is thick, cars back up in line to get to the tollbooths, and after paying their tolls, drivers lose time scrambling for position as the many lanes exiting the toll plaza merge together, returning the road to its original width. Its a real problem, as confirmed by the experience of transportation departments around the world. A study conducted at the New Jersey Institute of Technology estimates that a travel time savings of 2 minutes, or over 10 percent, could be affected by the removal of two toll plazas along a 14-mile section of the Garden State Parkway [10]. Modern toll facilities, such as Highway 407 near Toronto and the SR-91 Express Lanes in Orange County, California, require all payment to be made by means of electronic transponder, so that vehicles do not have to slow down in order to pay the toll [9]. But on many older tollways moving to all-electronic payment is not an option, while mounting congestion means that planners are faced with the problem of configuring their existing infrastructure to provide the best possible service.

To get rid of long lines, common sense suggests that providing as many tollbooths as possible will minimize the toll payment delay, but more tollbooths mean more merging and hence more congestion after the tolls are paid. In this paper we will investigate the question of choosing the right number of tollbooths to get an optimal balance between these two factors.

In doing so we will analyze the situation from a qualitative standpoint to determine what parts of the problem we should attempt to model, what we should ignore, and what rules the components of the system should be expected to follow, as well as establishing a metric for highway performance. Then we will be able to apply some results from queuing theory to derive a means of computing the performance level for a toll plaza given the number of initial lanes, number of tollbooths, and traffic flow level, from which we will be able to determine, fixing the other parameters, the number of tollbooths which performs the best.

2 The Toll Plaza

2.1 Types of Toll Plazas

Roadways on which tolls are collected are almost always so-called *controlled access* highways, which can only be entered and exited from designated access points, and are usually fully grade separated from other roads. There are three primary systems for collecting tolls on tolled highways. The interaction between the number of tollbooths and congestion is most obvious when the barrier toll system is used. In this case, vehicles may enter and exit the mainline roadway without paying a toll, but at certain locations along the highway there are toll plazas at which all vehicles must stop and pay a toll.

Another common possibility allows free flowing traffic on the mainline, but places tollbooths on either entrance or exit ramps, requiring tolls to be paid there. Where this arrangement is used, several factors are present which change the overall relationship between congestion and toll booth configuration. First, the flow of vehicles entering or exiting at any given interchange is usually much less than the flow on the mainline roadway, so that effects which only manifest themselves under heavy congestion may not ever occur. Second, to move between a limited-access toll freeway and an arterial surface street, it is often necessary, as well as queuing in line at the tollbooth, to pass through one or more traffic control devices such as stop signs, traffic signals, or roundabouts, which can have their own lines and produce additional delay not directly attributable to the toll collection system. The modeling of these effects does not contribute any insight to the central question of toll plaza delay, and reduces the likelihood of developing a computationally tractable model. When the entrance and exit ramps are heavily used and traffic on the ramps and between the ramps and the roadway they connect to otherwise flows freely, these toll plazas can be treated as equivalent to those in the barrier toll case, and our model will accommodate them without modification. We will not attempt to model the other cases discussed above. Therefore, we will not specifically consider entrance and exit toll booths subsequently.

The third major system, mentioned in the introduction, is electronic toll collection. In this system, drivers identify themselves with electronic transponders, and when they enter the highway the toll is automatically deducted from an account previously established. From a traffic flow standpoint this is no different from an ordinary freeway with no tollbooths, so any questions relating specifically to tollbooths do not apply. We therefore do not consider electronic toll collection in the remainder of this paper.

2.2 Approaching the Toll Plaza

We now examine the physical layout and operation of a toll plaza in the barrier toll case in greater detail. We assume that the highway is generally free flowing on either side of the area of the toll plaza. This assumption allows congestion resulting from the toll plaza design to be isolated from general congestion on the highway. The capacity of a highway is given in terms of the traffic flow rate: the number of vehicles passing a single point per time unit. Since we are assuming the roadway itself is free-flowing, we will only consider traffic flow levels through the toll plaza less than the capacity of the road. Empirical analysis of traffic

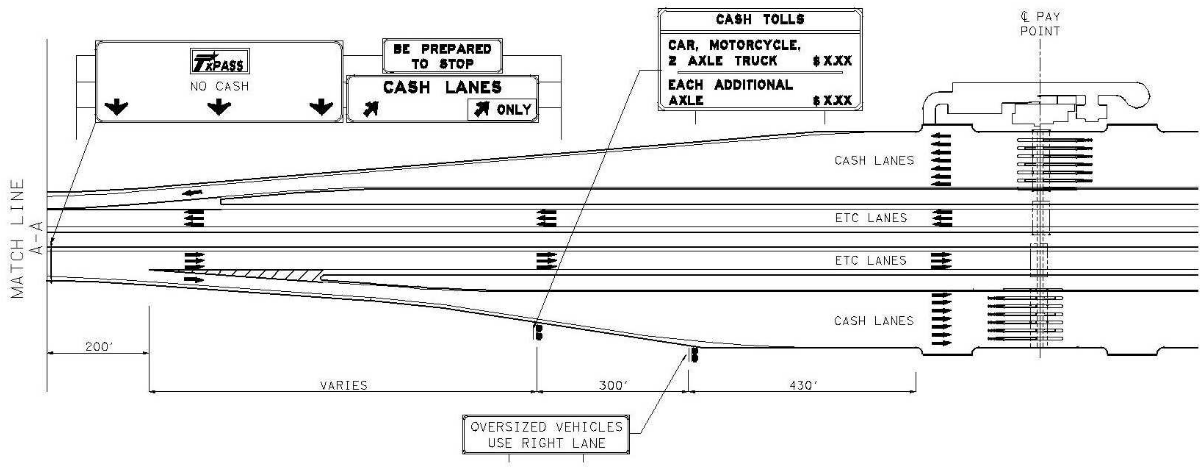

Figure 1: Concept design for an actual barrier toll plaza, showing queuing area and tollbooths, from Rastorfer [11]

levels on highways indicates a maximum capacity of approximately 2000 vehicles per hour on each lane [8, 3].

Traffic flow is usually considered to be roughly constant at any given instant, as changes in flow occur smoothly and slowly, while measurements employed are over very short time periods. This means that, viewed as a stochastic process, traffic inter-arrival times must follow an exponential distribution.

As one approaches the toll plaza from one direction, if the number of toll booths is greater than the normal number of lanes, the number of lanes increases to equal the number of toll booths, and signage instructs drivers to prepare to stop. At some point most vehicles must stop, either because the vehicle in front of them has stopped, or they have reached a toll booth. We assume that this *queuing area* where the roadway has widened to a number of lanes equal to the number of toll booths, and vehicles wait in line to pay tolls, is sufficiently long that all of the waiting lines, at any time, will be contained inside it, so that the waiting lines can be analyzed as a set of queues equal in number to the number of tollbooths, without concerning ourselves with vehicles that could spill outside the queuing area in heavy congestion. This assumption does not hold for all currently operating tollways, but every newly designed toll plaza should be capable of accommodating the volume of traffic it is expected to carry. We also assume that the amount of time a vehicle spends between entering the queuing area and stopping at the end of a tollbooth line is constant regardless of the lengths of the lines or the particular booth which the driver chose.

2.3 Tollbooths

In an actual toll plaza a variety of tollbooth types can be found. The slowest, and most general, tollbooths, labeled as "cash" lanes, allow payment to be made with cash, with the attendant making change if exact change is not paid. In "automatic" lanes, the driver must pay with exact change in order to minimize payment time, the booth being operated automatically. Finally, there are two types of lanes for electronic payment that can be present, even when this is not the only mode of payment. Some electronic payment lanes still require the driver to stop at a booth and wait for the payment to be made, usually a result of retrofitting electronic payment systems onto toll plazas not designed to accommodate them. There are also "express" lanes that allow drivers paying electronically to pass through the plaza at full highway speed [2]. Since vehicles on the express lanes do not have to stop, they are not relevant to the problem at hand, so we will consider toll plazas without express lanes. The other types of booth, although they operate at different speeds, all have essentially the same behavior.

The question of determining the optimal mix of cash, automatic, and non-express electronic booths is another optimization problem which a tollway operator employing all of those types of booths would have to solve, but as it is tangential to the issue of congestion due to queuing and merging, and any solution to it requires knowledge of the distribution of drivers choosing to pay by each method, we will not attempt to solve it. Instead we will suppose all of the toll booths in our plaza are identical, and the average time to pay the toll at any of them is equal. The maximum vehicle flow per toll booth has been observed to lie between 350 and 500 vehicles per hour for standard (cash and automatic) tollbooths [2, 3]. Like traffic flow on the highway, this rate can also be considered to be rougly constant in local time intervals, so the toll payment time should also follow an exponential distribution.

2.4 Merging

After the tollbooths, the roadway must narrow back from a number of lanes equal to the number of tollbooths, to its normal width, a section we will call the *merging area*. How abruptly this happens varies tremendously in actual practice. Sometimes the extra lanes end almost immediately, forcing a sharp merge at a relatively low speed. In other cases the additional lanes extend far enough for drivers to reach full highway speed before they are required to merge. We will generally suppose that, as a newly-designed toll plaza with minimal spatial constraints, our plaza's merging area is sufficiently long to allow all vehicles to reach highway speed before merging. Additionally, different merging patterns are used when lanes begin and end. With several lanes merging into one, all of the merging could occur at a single point, but this means that as many vehicles as there are lanes could interfere with each other at that point. For a smoother transition solutions involving only the merging of pairs of lanes are used. One common choice is to always merge out the rightmost (or leftmost) lane until the desired number of lanes is reached. This pattern is very advantageous for the side of the roadway where no merging occurs, but drivers on the other side could be required to merge many times. Another possibility is a "balanced" pattern where pairs of adjacent lanes all across the roadway merge repeatedly until the desired roadway width has been attained. This distributes the merges more evenly over the roadway. We will generally assume lanes merge out on one side, two at a time.

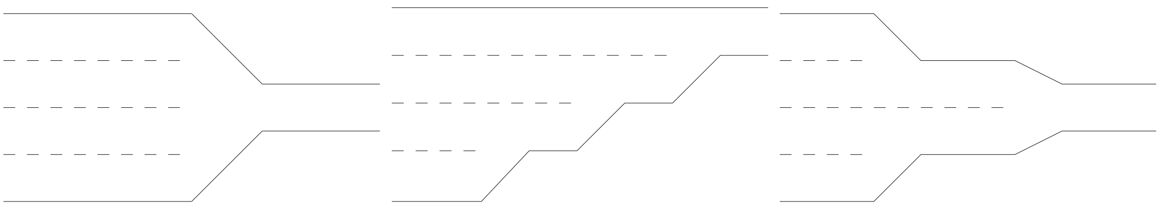

Figure 2: Four lanes merging to 1 at a single point (left); with rightmost lanes ending (center); in a balanced pattern (right)

3 The Driver

3.1 Irrational Drivers

The actual behavior of drivers in the vicinity of a toll plaza can be unpredictable and lead to worse than expected performance of the entire facility in some cases.

Drivers may choose a tollbooth that already has a long line over a shorter one, increasing the time they ultimately spend in the facility. They may even choose an occupied lane when an empty lane exists. In an observational experiment by Leslie Edie at the Lincoln Tunnel, "the number of instances when there were one or more lanes empty in the example was 31 out of 40, thus giving a percentage availability of 77.5" [4].

Drivers can perform dangerous maneuvers that lead to collisions and ultimately the closure of a portion of the facility, reducing its capacity. They can enter a booth without enough money to pay the toll, holding up traffic as they argue with the attendant and, possibly, are forcibly removed from the booth. Undeniably, poor driver performance leads to decreased toll plaza performance, but should this be accounted for in a toll plaza model, and if so, how? Modeling irrational driver performance is difficult, because by their very nature the things bad drivers do are not easily predictable. A model that merely assumed drivers would perform one of a predefined list of bad behaviors with a certain probability could not completely describe every mistake that might be made in the course of navigating a toll plaza. If it was necessary to accurately model irrational driver behavior it could therefore be incredibly complicated.

When choosing a driver model we must consider that we are, in fact, looking for a global optimum solution. Given that any reasonable model should perform better with better drivers, the absolute best performance should be obtained with a model of completely competent drivers, along with a toll plaza configuration that yields the best performance with competent drivers. Therefore, we will assume all drivers in our system are both rational and competent. We will now attempt to describe the behavior of such a driver in the toll plaza described in the last section.

3.2 Rational Drivers

Most users of transportation facilities are not using them for their own sake, but rather as a means from getting from one place to another, and will either be, for reasons of schedule and fuel cost, interested in moving as quickly as possible, or at least not disadvantaged by shorter travel times versus longer ones. Therefore, we assume a rational driver's objective in the toll plaza is to minimize their own travel time, subject to constraints such as traffic laws, other vehicles, and their own vehicle's performance. As we have argued, there are two things that determine the waiting time in the toll plaza; the queue length and the congestion due to merging after the toll booths. A driver entering the queuing area has no immediate knowledge of the conditions in the merging area, and so a rational driver will believe the likelihood of encountering merging congestion in each lane is equal. They will not choose any lane over another, therefore, on the basis of merging considerations. The driver will, however, be aware of the lengths of the lines ahead of the different toll booths, and know that, as the time to pay the toll in each booth is roughly equal, getting into a shorter line will result in a shorter wait time. A rational driver must, therefore, always choose to enter the shortest line in the merging area. This leads, at a large scale, to entering vehicles being evenly distributed among the available tollbooth lines.

Having paid their toll in an orderly and efficient manner, the rational and competent driver will apply standard safe driving practices when exiting the merging area. Based on the time it takes for a vehicle to slow to a stop, and the time it takes for a person to react to visual stimulus, it is recommended that drivers follow any leading car by no less than 3.5 seconds. Because it takes in the vicinity of 10 seconds to pay the toll under our model, it is unlikely that a vehicle leaving the same tollbooth as our driver, and in the same lane in the merging area before a merge point, will not be a safe following distance in front, so the driver will move unobstructed to near the merge point. Here, however, if another vehicle coming from the other lane has already passed the merge point, they will be in front of the driver's car, and if they are less than 3.5 seconds ahead of the driver, the driver must slow down and delay merging until the other car is more than 3.5 seconds ahead. If there is no car within the safe following window the driver may continue through the merge point at full speed.

4 The Utility Metric

4.1 Measuring Toll Plaza Performance

So far, we have discussed the performance of a toll plaza, without defining exactly how performance should be measured. As we suggested in the previous section, the objective of a driver on any road is to minimize travel time. It might seem intuitive, therefore, that the object of a toll highway should be to minimize the average travel time of all drivers on that road. There are, however, other aspects to highway performance that are frequently considered by transportation engineers.

4.2 Environmental Impact

Current law in many countries require the environmental impact of new highway construction to be analyzed prior to the start of construction, and the design to have as little impact on the environment as possible. The most obvious environmental impact of a highway is pollution due to vehicle exhaust. Noise from freeways is also a major concern in many communities, as is the displacement of wildlife habitat. It does not seem reasonable to take one of these alone as the sole quantity to optimize, and completely ignore travel time. After all, the primary users of any highway facility are those who use it for travel, and their primary interest is optimizing travel time. So if we do consider environmental factors, we must optimize them in conjunction with travel time, which means that we must find a single utility function depending on travel time and environmental factors, combining them in a manner which reflects the actual combined value placed on them by the entire community.

This seems like an insurmountable problem, except when the nature of the environmental factors is considered. Wildlife habitat is destroyed through the mere existence of the highway, and much of the damage occurs due to the action of the road as a barrier (regardless of its width) and the other pollution and noise effects caused by automobiles on the road. The latter does not depend on the number of tollbooths, so it is impossible to optimize, and we are left with pollution and noise only. However, pollution and noise both exist in direct proportion to the number of automobiles in the area of the toll plaza at any time. So minimizing environmental effects is best achieved through reducing the number of automobiles, and as we have no control over incoming traffic flow, the only way to do this is by reducing travel time through the toll plaza. To first order, we can achieve the lowest environmental impact through minimizing travel time. Therefore, we will not specifically consider environmental impact any further.

4.3 Traffic Capacity

Another quantity which is important in relation to roads is the traffic capacity and traffic flow through the toll booth. We might be interested in maximizing these as well as travel time, as it is clearly in the interest of the tollway operators and the public for the highway to move as many people as possible. However, the traffic flow into the toll plaza is fixed by other parts of the highway, so we cannot change the flow through the plaza except by constricting it with an inadequate number of toll booths. Any reasonable solution, including that which minimizes expected travel time, should have sufficient capacity to carry the incoming traffic flow. However, increasing capacity beyond this will not lead to an increase in the overall capacity of the highway, or increase the total traffic flow; it will only shift the bottleneck outside of the toll plaza. Since there can be no benefit to specifically addressing capacity or traffic flow, we will ignore it in determining optimality.

At this point we can define an optimal toll plaza configuration: It is the one which minimizes the expected time a rational driver must spend traveling through the system.

5 Assumptions

Here we summarize some aspects of our toll booth construction, based on the preceding discussion, which will be employed in the development of our model.

- **The traffic flow is constant in a short period.**
- **The time between two cars entering the toll plaza is of exponential distribution.**

Table 1: Notation of Queueing system. Information from [6].

Characteristics	Symbol	Description
Arrival pattern(A)	M	Exponential distribution
Service pattern(B)	M	Exponential distribution
	G	General distribution
Number of servers(X)	$1, 2, \ldots, \infty$	
System capacity(Y)	$1, 2, \ldots, \infty$	
Queue discipline(Z)	FCFS	First come, first served

- **The traffic streams fan out into tollbooths smoothly and evenly.** We have assumed the "fan-out" into the toll booth does not contribute to delay in a manner which is dependent on the parameters under consideration.
- **The drivers are delayed by waiting in lines for toll collection.** If there is a line at the chosen tollbooth when the driver arrives, the driver will have to wait until other cars have left to enter the booth.
- **The drivers are delayed by toll collection, and the delay is distributed exponentially.**
- **The drivers are delayed by the merging process after leaving the tollbooths.**
- **The toll plaza adopts the side merging layout at the exit.**

6 Queueing Model

Under our hypotheses, traffic delay in a toll plaza is caused by toll collection and car stream merging. Therefore, we break the problem into two parts: delay in tollbooths and at merging points. In a tollbooth, drivers wait for the service of toll collection, while at a merging point, drivers may stop to wait for a chance to get onto the merged lane. Here we adapt Queueing Theory and see each part as a queueing system.

Queueing Theory is a theory that models the process when customers line up to wait for service. A queueing system has the following basic characteristics: arrival pattern of customers, service pattern of servers, number of servers, system capacity, and queue discipline. The standard notation for describing the configuration of a queueing process is **A/B/X/Y/Z**, where the legend is given by Table 1. (In shorthand, system capacity, Y, and queue discipline, Z, can be omitted when the capacity is infinite and the queue discipline is first-come first-served.)

6.1 Tollbooth

Under our assumptions, the traffic streams coming from the entrance of the toll plaza are evenly fanned out into all tollbooths and each tollbooth will recieve a stream whose inter-arrival time is of exponential distribution. In addition, the service time also has an exponential distribution. Thus, we can see each tollbooth as an independent **M/M/1** queue.

Burke's Theorem provides that the outcoming stream from the tollbooth also has the exponential distribution with the same rate as the arrival stream [7]. We will use this property in the queue model of merging points.

6.2 Merging Point

The total delay by the entire merging process is more complicated to analyze. We shall first consider the simple merging process when cars from two lanes merge into one, called **2-into-1 merging point**. When a driver on one lane arrives at the merging point, the delay time depends on whether there is another car on the other lane. If the other lane is empty, the driver can directly pass through, otherwise he or she has to stop and wait.

For simplification, we treat the two incoming lanes as one queue. Since we cannot distinguish between cars originating in either lane, the driver now must stop and wait whenever there is another car in the queue. We define the service time of a car as the time it spends to pass through the merging area. Under this definition, the service rate is equal to μ_B when more than one car is in the system, and μ_0 when the system has only one car, where μ_0 and μ_B are constants representing the waiting time when a driver does or does not have to yield to another car when merging.

Therefore, the service pattern of this queueing system is a general funtion. If the arrival pattern is exponential, we can set the configuration of the system as **M/G/1**.

6.3 Total Merging Process

We consider the total merging process in the plaza as multiple 2-into-1 merging points. If there are T tollbooths and afterward the streams are merged back into N lanes, the total number of merging points would equal $T-N$. However, their arrival rates are different. The arrival rate of a merging point equals the traffic flow it recieves. If a merging point recieves a traffic stream coming from k tollbooths, having a total traffic flow Φ, its arrival rate would be:

$$\lambda = \frac{k}{T} \times \Phi,$$

The values of k for merging points depend on the merging layout. For example, in a toll plaza with a side merging layout which has T tollbooths, the first merging point takes a stream coming from 2 tollbooths, and the second merging point would take a stream from 3 tollbooths, etc.

The overall average wasted time is the weighted sum of all averaged wasted time at each merging point, where the corresponding weight is the probability for a driver to reach that point, which is k/T. Suppose a toll plaza has T tollbooths, N lanes at the exit, and recieves

Table 2: Arrival rate and probability at each merging point.

Merging point	1_{st}	2_{nd}	3_{rd}	...	$(T-N)_{th}$
Arrival rate	$2\Phi/T$	$3\Phi/T$	$4\Phi/T$	...	$(T-N+1)\Phi/T$
Probability	$2/T$	$3/T$	$4/T$	...	$(T-N+1)/T$

total traffic flow Φ. Then, the arrival rate and corresponding probability at each merging point is shown by Table 2.

7 Calculation

Having the model ready, we can calculate the formula for wasted time in terms of the number of tollbooths. However, since our model relies on the values of some constants, we must estimate these first.

7.1 Estimates of Constants

- **Number of incoming lanes** (N)
  The typical range of number of lanes on a highway (in one direction) is about 1 to 6. We will consider different values of N in our calculation.

- **Total traffic flow** (Φ)
  The maximum traffic flow per lane is 2000 (1/hr) [8]. While N ranges from 1 to 6, we would consider various values for traffic flows, including heavy and light traffic conditions.

- **Service rate at a tollbooth** (μ_A)
  The service rate of a tollbooth is about 350 vehicles per hour [2, 3].

- **Service rate at a merging point: when merging does not occur** (μ_0)
  μ_0 is the service rate when there is only one car in the merging point system. This value is the time for a car to drive through the merging area at the average highway speed. The average highway speed is 60 mph [8]. The length of the merging area is the average car length plus a safety distance, which is around $15 + 6 \times 15 = 105$ (ft) [8]. Thus the average service time here is 105 ft/60 mph $= 1.1932$ sec, and the service rate $\mu_0 = 3600/1.1932 = 3017.1$ (1/hr).

- **Service rate at a merging point: when merging occurs** (μ_B)
  To estimate μ_B we need to consider the time a vehicle passes through the same area with zero initial speed. Under this speed the safety distance would be one car length, and the average acceleration of a vehicle is 6.5 [8]. Therefore the average serive time is $(2 \times (15+15)/6.5)^{1/2} = 3.0382$ (sec), and the service rate $\mu_B = 3600/3.0382 = 1184.9$ (1/hr).

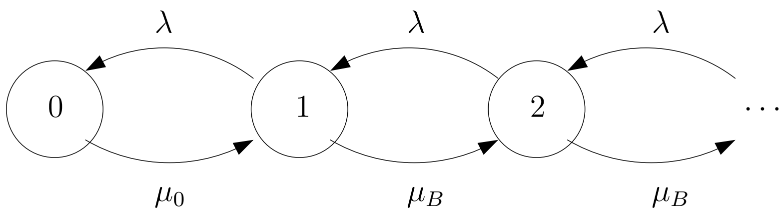

Figure 3: State Diagram of Merging Point Queue

7.2 Wasted Time - Tollbooth

Given the above definitions the arrival rate of each tollbooth is Φ/T. From the arrival rate and service rate of the tollbooth, according to the performance measure formula [7], the average service time of each tollbooth is

$$w_A = \frac{1}{\mu_A - \Phi/T}$$

7.3 Wasted Time - Merging Point

A merging point is modeled as a Markovian system as depicted in Figure 3, where each state represents the number of vehicles in the system. Notice the single μ_0 in Figure 3 which represents the possibility of no merging conflicts. The arrival rate is some value λ. We will calculate the average wasted time $t_{diff}(\lambda)$ after first calculating the average waiting time in the system $t_{sys}(\lambda)$.

Let P_n be the probability that there are n drivers in the system. When the system reaches equilibrium, the net probabilty of trasition is zero for each state. In addition, the sum of all P_n must be one. Therefore we have

$$\begin{cases} \lambda P_0 = \mu_0 P_1 \\ \lambda P_1 + \mu_0 P_1 = \lambda P_0 + \mu_B P_2 \\ \lambda P_n + \mu_B P_n = \lambda P_{n-1} + \mu_B P_{n+1}, n \geq 2 \\ \sum_0^\infty P_i = 1 \end{cases}$$

Solving the equations we obtain

$$\begin{cases} P_0 = (1 + \frac{\lambda}{\mu_0} + \frac{2\lambda^2 \mu_B}{\mu_0(\mu_0+\lambda)(\mu_B-\lambda)})^{-1} \\ P_1 = \frac{\lambda}{\mu_0} P_0 \\ P_n = \frac{2\lambda^2}{\mu_0(\mu_0+\lambda)} (\frac{\lambda}{\mu_B})^{n-2} P_0, n \geq 2 \end{cases}$$

And we can calculate the expected number of drivers in the system:

$$L(\lambda) = \sum_0^\infty i P_i = \frac{\lambda}{\mu_B - \lambda} + \frac{\lambda(\mu_B - \mu_0)}{\lambda(\mu_B - \mu_0) + \mu_0 \mu_B}$$

By Little's Theorem, the average waiting time in the system $t_{sys}(\lambda)$ equals the expected number of drivers in the system $L(\lambda)$ divided by the arrival rate λ (see [7]).

$$t_{sys}(\lambda) = \frac{L(\lambda)}{\lambda} = \frac{1}{\mu_B - \lambda} + \frac{\mu_B - \mu_0}{\lambda(\mu_B - \mu_0) + \mu_0\mu_B}$$

The average wasted time of a driver at a merging point is the difference between t_{sys} and the time he or she would spend on a normal lane. The expected time a driver spends when no merging happens is $\frac{1}{\mu_0}$. Hence,

$$t_{diff}(\lambda) = t_{sys}(\lambda) - \frac{1}{\mu_0} = \frac{1}{\mu_B - \lambda} + \frac{\mu_B - \mu_0}{\lambda(\mu_B - \mu_0) + \mu_0\mu_B} - \frac{1}{\mu_0}$$

7.4 Wasted Time - Total Merging Process

According to the discussion in the previous section, under a side merging layout, the arrival rates of the merging points are:

$$\frac{2\Phi}{T}, \frac{3\Phi}{T}, \frac{4\Phi}{T}, \ldots, \frac{(T-N+1)\Phi}{T}$$

The corresponding probabilities of reaching these merging points are:

$$\frac{2}{T}, \frac{3}{T}, \frac{4}{T}, \ldots, \frac{T-N+1}{T}$$

The overall wasted time is the weighted sum:

$$w_B = \sum_{i=1}^{T-N} \frac{i+1}{T} \times t_{diff}\left(\frac{i+1}{T} \times \Phi\right)$$

Since each merging point has a different arrival rate and t_{diff} is complicated, it is difficult to find a closed form for the sum. However, it is simple for a program to compute this value. Finally, the average wasted time a driver spends in the entire toll plaza is the sum of w_A and w_B.

$$w_{total} = w_A + w_B = \frac{1}{\mu_A - \Phi/T} + \sum_{i=1}^{T-N} \frac{i+1}{T} \times t_{diff}\left(\frac{i+1}{T} \times \Phi\right)$$

Note that this wasted time value represents the portion of the total travel time through the toll plaza which, in our model, depends on the parameters being varied, so by minimizing this we are also minimizing the total travel time.

8 Results

Based on the formulas given in the previous section, the average wasted time w_{total} is determined by the number of lanes N, number of tollbooths T, and traffic flow Φ. We use Matlab to compute w_{total} for various N, T, Φ. Figure 4 shows the average wasted time for different value of T when $N = 1$ and $\Phi/N = 900$.

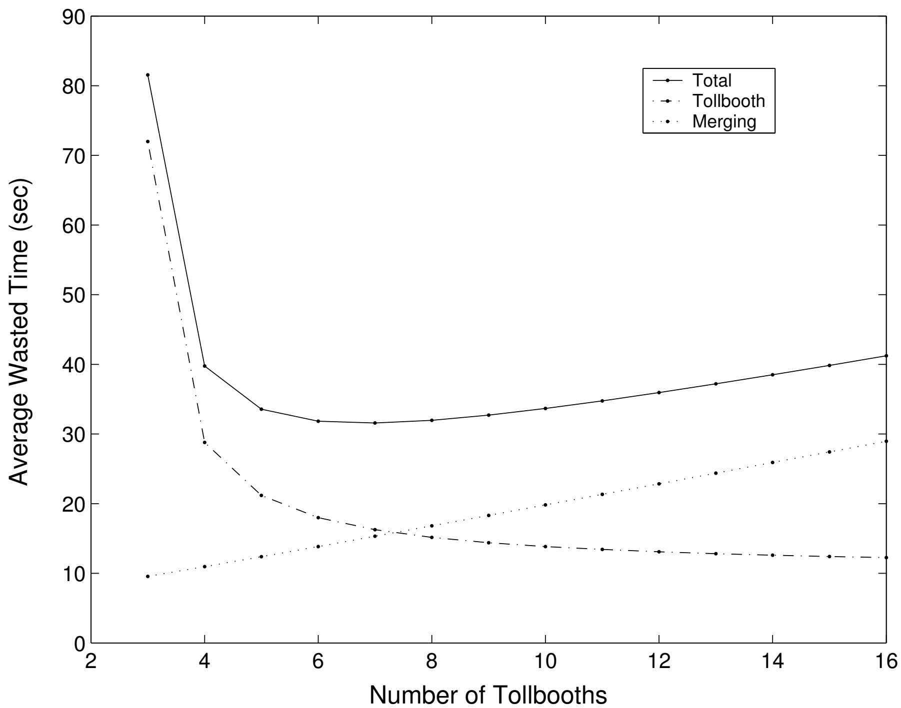

Figure 4: Average Wasted Time when $N = 1$, $\Phi/N = 900$.

The minimum value of total average wasted time is 31.6 sec, which happens when $T = 7$, so 7 tollbooths is optimal for this configuration. The curve of total wasted time is composed by the curves of time wasted in tollbooth and merging. These two curves indicate that as the number of tollbooths increases, *the tollbooth delay decreases almost harmonically, while the merging delay increases.*

We pick out the optimal T values for many pairs of N and Φ (Figure 5, Table 4, Table 5). However, an optimal T is not guaranteed for each possible configuration. *When the traffic flow is too large, every value of T results in an overflow situation* (which means either a tollbooth or a merging point cannot hold its arrival rate), so no optimal value of T can be found.

9 Conclusion

9.1 The "One-Tollbooth-per-Lane" Scenario

With the standard setting of our model, the toll plaza works inefficiently when the number of tollbooths and the number of incoming lanes are the same. The benefit of adding a first tollbooth always overcomes the disadvantage of the marginal merging process. However, if the toll plaza has a high service rate of toll collection (for example, an electronic payment system) so that merging becomes the main reason of congestion, this scenario would be the best choice. Table 3 shows that as the service rate μ_A at toolbooths increases, the optimal number of tollbooths gets closer to the number of lanes N.

Table 3: Optimal Number of Tollbooths with Different Service Rates

Service Rate	$T = 4$	$T = 5$	$T = 6$	$T = 7$
300	–	–	–	301.1
500	–	38.0	29.8	65.9
700	18.0	14.0	18.0	57.8
900	9.0	9.2	14.6	54.9
1100	6.0	7.1	12.9	53.5
1300	4.5	6.0	12.0	52.6

9.2 Recommendation

Based on the discussion above, we recommend to use one tollbooth per incoming lane when the service rate of toll collection is very high. Otherwise, we suggest to choose the optimal number of tollbooths from Table 4, depending on the traffic condition.

References

[1] Agboh, Dennis K. and Anthony A. Saka, *Assessment of the Impact of Electronic Toll Collection on Mobile Emissions in the Baltimore Area*, National Transportation Center, Baltimore, 2000.

[2] Al-Deek, Haitham M. and Jack Klodzinski, "Proposed Level of Service Methodology for Toll Plazas", presentation at the Transportation Research Board Annual Meeting, National Research Council, Washington, 2002.

[3] Ceballos, Gustavo and Owen Curtis, "Queue Analysis at Toll and Parking Exit Plazas: A Comparison between Multi-server Queuing Models and Traffic Simulation". `http://www.ptvamerica.com/docs/VISSIMQueueAnalysis.pdf`

[4] Edie, L. C., "Traffic delays at toll booths", *Journal of Operations Research Society of America* 2 (1954) 107-138.

[5] Delaware Department of Transportation, "Dedication Ceremony Held for Opening of E-ZPass Express Lanes at the Dover Toll Plaza". `http://www.deldot.net/public.ejs?command=PublicNewsDisplay&id=1823&month=5&year=2004`.

[6] Gross, Donald and Carl M. Harris, *Fundamentals of Queueing Theory*, Wiley, New York, 1998.

[7] Hock, Ng Chee, *Queueing Modeling Fundamentals*, Wiley, New York, 1996.

[8] Institute of Transportation Engineers, *Traffic Engineering Handbook*, Ed. James L. Pline, Prentice Hall, Englewood Cliffs, NJ, 1992.

[9] Intelligent Transportation Systems Society of Canada, "ITS Canada: Electronic Toll Collection". http://www.itscanada.ca/english/2reps2e.htm

[10] New Jersey Institute of Technology - National Center for Transportation and Industrial Productivity / International Intermodal Transportation Center, "Ten Year Plan to Remove the Toll Barriers on the Garden State Parkway". http://transportation.njit.edu/NCTIP/final_report/GSPFinalReport.htm

[11] Rastorfer, Robert L., "Toll Plaza Concepts", presentation at ASCE Fall Conference, Houston, 2004.

[12] Wikipedia, "Toll road". http://en.wikipedia.org/wiki/Toll_road

A Data and Graphs

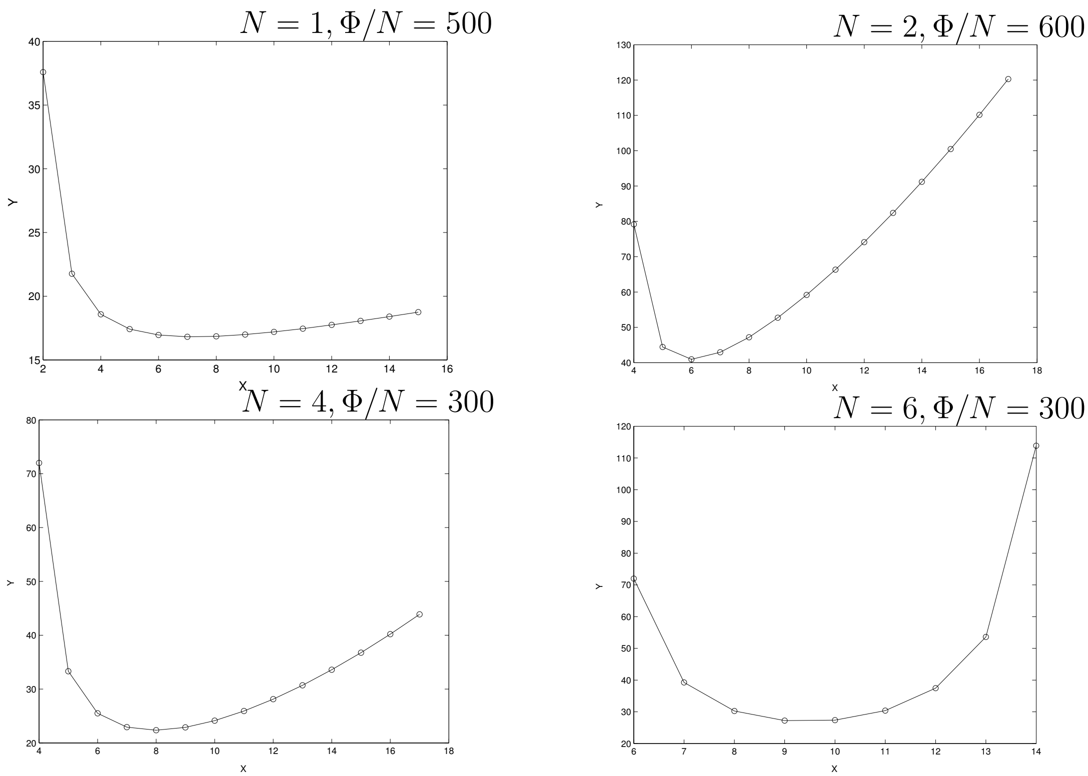

Figure 5: Average Wasted Time under Various Configurations.

Table 4: Optimal Number of Tollbooths

Flow per Incoming Lane (1/hr)	Number of Incoming Lanes: 1	2	3	4	5	6
100	7	7	8	8	9	10
200	7	7	8	8	9	10
300	7	7	7	8	9	9
400	7	7	7	7	9	9
500	7	7	6	7	–	–
600	7	6	–	–	–	–
700	7	5	–	–	–	–
800	7	–	–	–	–	–
900	7	–	–	–	–	–
1000	7	–	–	–	–	–
1100	6	–	–	–	–	–
1200	–	–	–	–	–	–

Table 5: Average Wasted Time (sec)

Flow per Incoming Lane (1/hr)	Number of Incoming Lanes: 1	2	3	4	5	6
100	11.2	11.9	12.4	12.7	12.9	13.1
200	12.3	14.0	15.3	16.3	17.1	17.8
300	13.6	16.7	19.7	22.4	24.9	30.3
400	15.0	20.7	27.6	36.3	45.2	56.5
500	16.8	27.4	48.2	105.1	–	–
600	19.0	40.9	–	–	–	–
700	21.9	95.8	–	–	–	–
800	25.8	–	–	–	–	–
900	31.6	–	–	–	–	–
1000	41.8	–	–	–	–	–
1100	69.7	–	–	–	–	–
1200	–	–	–	–	–	–