

Queue Analysis at Toll and Parking Exit Plazas: A Comparison between Multi-server Queuing Models and Traffic Simulation

Gustavo Ceballos and Owen Curtis

Abstract. Multi-server queuing analysis can be used to estimate average wait times and queue lengths at toll and parking exit plazas given arrival rates, number of servers, and service rates. These queuing models approximate the performance of queuing systems with multiple queues. While the queuing equations are useful where there are multiple servers, they do not fully reflect the operation of a toll plaza where vehicles can stack in multiple queues and can change lanes. To gain an understanding of how such a system will operate, a traffic simulation model that reflects the specified vehicle arrival and service rates, and allows for multiple channel arrivals was applied to several parking exit plaza concepts. This paper compares the results obtained from both methods and describes how these estimations compared with queue and waiting time data collected at exit plazas in actual operation.

The analysis here described was based on the application of the analytical queuing model in conjunction with traffic simulation in parking exit plazas at four different airports: the Ronald Regan Washington National Airport, the Dulles International Airport, the Salt Lake City International Airport, and the Atlanta Hartsfield International Airport. The measures of effectiveness used to evaluate the models' performance were average and maximum queuing delay, and average and maximum queuing length. Additionally, the output values, including the throughput volume estimated by the simulation model were compared to collected field values.

The paper shows that simulation provides a more comprehensive understanding of the toll plaza operation allowing for a more in-depth analysis of its performance. While the simple queuing models could be used as an initial, in some cases more conservative estimation in early stages of planning, simulation should be used for advance planning, design, operation and management of toll and exit plaza facilities.

1. INTRODUCTION

Highway toll plazas and parking entry/exit plazas constitute a unique type of transportation system that requires special analysis when trying to understand their operation and their interaction with other roadway components. On the one hand, these facilities are one of the most effective means of collecting user fees either for parking services or for roadways, bridges, and tunnels. On the other hand, toll plazas adversely affect the throughput or capacity of the facilities they serve. The adverse effect of toll plazas is particularly evident during hours when traffic is usually heavy. When each vehicle must come to a stop in order to be processed, the throughput per lane drops from freeway rates of 2,000 vphpl to toll plaza rates in the range of 350 to 500 vehicles per hour (vph). By contrast, service rates for parking entry plazas range up

to 660 vph, while service rates for parking exit plazas range from 80 to 250 vph depending on type of payment. Based on these numbers, it is not surprising that both highway toll and parking plazas experience lengthy vehicular queues and long delays when demand is near or exceeds processing capacity. Efficient sizing of toll plazas becomes critical in minimizing the space requirements and capital expense of collecting user fees.

While the authors recognize the significant differences that exist between the operation of parking exit plazas and that of toll plazas on toll roadways, bridges, and tunnels, for simplicity reasons, all fee plazas are called hereto “toll plazas” or “toll facilities”. Section 5 explains in more detail some of these differences and the implication related to the analysis here presented.

Performance evaluation of toll facilities requires a good understanding of their unique characteristics and constraints, and the definition of the right measures of effectiveness (MOEs) that help explain the level of service perceived by users. The length of the queue and the time spent in the queue are perhaps the two most significant MOEs that reflect both how the system is performing and what users perceive as its level of service. Therefore, the right set of tools and techniques are necessary in order to produce these MOEs.

The analysis of queuing systems and its variables has been the focus of many studies and researches for many decades. Several mathematical models have evolved in an attempt to analyze different queuing, dispatching and processing conditions. It is not the purpose of this paper to explain the mathematical foundation and the scope of each of these models, other than to describe two of them -- the multi-server queuing and the multiple single-server queue models - how they have been applied in planning and evaluation of toll plazas, and how they compare with other available techniques such as stochastic simulation models. As explained below, toll plazas can be treated, with some limitation, as either single queue, multi-server or as multiple single-server queuing systems in order to determine queue length, user waiting time, and other statistics. Even though a more accurate representation of multiple queuing at toll plazas may be obtained by applying a multi-queue, multi-server model, the analytical formulation for this type of model is extremely complex and furthermore, no single set of equations are available to estimate its parameters, requiring the use of complementary techniques such as Markov's processes¹ and discrete simulation.

Toll plazas have been the focus of simulation efforts for some time. Many of the efforts in toll plaza simulation have been event-discrete models. However, while these models are very well suited to represent the user-server process, in general they do not incorporate the series of complex traffic interactions and driver behaviors that take place on these facilities. Both the approaching area where vehicles decelerate and maneuver to get into the “most convenient lane” and the exit area where vehicles accelerate and must merge into fewer roadway lanes, are subject to weaving and unique right-of-way conditions (priority rules). These traffic characteristics have a significant impact on the overall operation of the plazas and on the perceived level of service. To account for these issues, new and more sophisticated simulation models have been developed in recent years that allow for a stochastic representation of both the server-user process and the traffic operations in the same model.

The VISSIM traffic simulation model was applied in the planning and evaluation of parking plazas described on this paper. VISSIM is a microscopic, time-step and behavior-based simulation model developed to analyze a full range of traffic operations on virtually any kind of roadway. Even though the model was not specifically developed for toll plazas, some of the software capabilities allow for the simulation of these type of facilities as an integral part of the roadway network being served by the plaza. The model is also very well suited to simulate complex vehicle interaction and different driver's behavior.

While simulation models are tools that are more sophisticated and allow for a more in-depth analysis, it is also true that they normally require more up-front time for development and production of results than mathematical models do. It is then of importance to understand how comparable results from both tools are, and to what extent they can be used interchangeably, or what is the range of applicability of each one.

2. DESCRIPTION OF THE MODELS

2.1 Multi-server Queuing Systems - M/M/s

The multi-server queuing mathematical model is known in Kendall's notationⁱⁱ as the M/M/s model, where:

- M signifies a Poisson distribution (see explanation below)
- s = number of parallel service channels in the system.

The M/M/s model is one of the most commonly used to analyze the queuing problem in toll plazas. This model computes average wait times and queue lengths, given arrival rates, number of servers, and service rates. For $s > 1$, the mathematical model is complex. When $s = 1$, a more readily calculable set of equations applies. This particular model applies where there are multiple channels served by a single queue, as at a bank teller or many airline ticket counters. The outputs of the model are:

- Expected waiting time per customer in the system
- Expected waiting time of customer in the queue
- Expected number of customers in the system
- Expected number of customers in the queue

The exact calculation of these measures requires knowledge of the probability distribution of the arrival rate and service time. Furthermore, even with that knowledge, the resulting formulae are exceedingly complex. Thus, some simplifying assumptions are required.

- The most basic of these assumptions is that the arrival rate obeys *Poisson* distribution, which is equivalent to saying that the interarrival times are exponential.
- The second assumption is regarding the nature of the probability distribution of the service times. With a *Poisson* distribution, the service times are assumed to be exponentially distributed.

Moreover, successive interarrival times and service times are assumed to be statistically independent of each other. Collectively, the Poisson assumptions of the M/M/s model make for a reasonably tractable solution.

Figure 1 below shows a generalization of the simple model of a multi-server queuing system. In this case, there are multiple servers, all sharing a common waiting line. If a user arrives and at least one server is available, then the user is immediately dispatched to that server. If all the servers are busy, a waiting line begins to form. As soon as one server becomes free, a user is dispatched from the waiting line using the dispatching discipline in force.

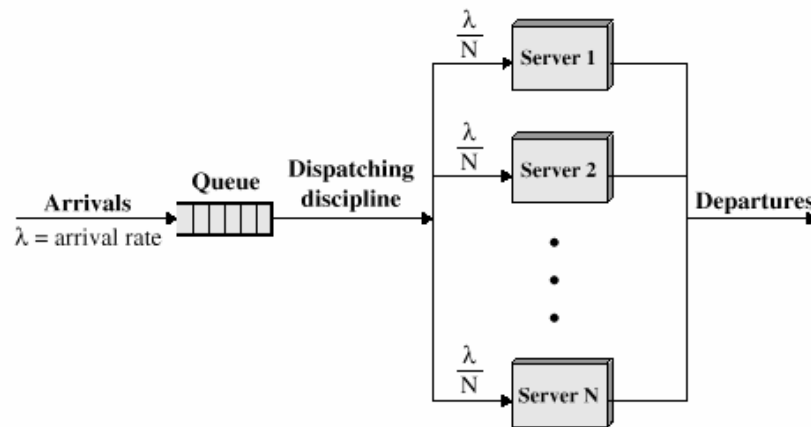


Figure 1: Multi-server Queuing Systems

The main equations that allow obtaining queue length and wait time are as follows:

$$L_q = \frac{\rho^{s+1}}{(s-1)!(s-\rho)^2} P_0$$

$$W_q = \frac{L_q}{\lambda}$$

where:

μ = mean service rate per busy server

λ = mean arrival rate

s = number of parallel servers

$\rho = \lambda \backslash \mu$ (traffic intensity = arrival rate / service rate, analogous to v/c in traffic operational analysis)

P_0 = probability of the system being empty

L_q = expected number of users in the queue

W_q = expected time spent waiting in queue

The computations associated with this model may be tedious. Therefore, in order to expedite the application of the model, a computer program was developed using the programming capabilities of standard spreadsheet software.

2.1.1 Limitations of this Model

When applied to the bank or airline ticket queue situations, this model can produce reasonable first-cut estimates, except in the cases where the arrival rate exceeds the collective processing rate. The M/M/s model cannot compute queue length and wait time when the number of customers arriving to be served is greater than the processing capacity (number of servers \times mean service rate = $\mu \times s$), in which cases the queue tends to infinity.

In a multiple-queue situation such as toll or parking exit plazas, there is no single queue. As explained before, this would require the far more complex calculations of a multi-queue, multi-server model, which this paper does not address. The difference between the simplified version and the reality of multi-queue plazas is in the ability to reflect complex driver behavior in terms of identifying the lane desired (based on queue length or other inputs which may indicate the path of fastest processing time) and the ability to do something about it (i.e., change lanes to get to the desired lane). In general, the application of a single-queue, multi-server equation would be expected to yield more favorable results than would occur with a multi-queue estimation.

2.2. Multiple Single-server Queuing Model

To work around the limitations of single queue M/M/s model for the applications of interest, an alternative approach is to model the multi-server queue as a series of single-server queuing systems (M/M/1) in parallel. Each of the single-server models serves λ/n users, where λ is the overall arrival rate and “n” is the number of single-server units. Figure 2 shows the generalization of this model. The same equations as the one mentioned above apply for each of the sub-models.

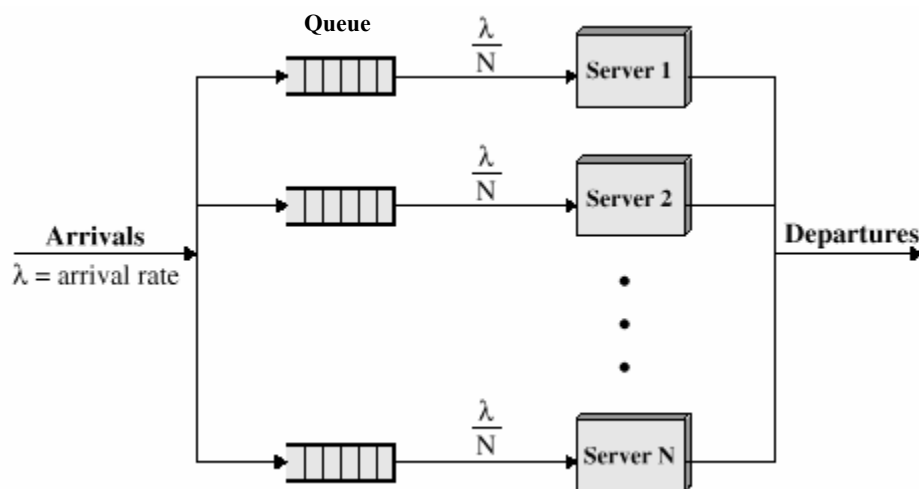


Figure 2: Multiple Single-server Queuing Systems

This application of the multiple single-server model may be used to overcome some of the limitations of the M/M/s systems mentioned above. Specifically, by segregating the system into “n” sub-systems, it accounts for some of the inefficiencies associated with a multi-queue system. In other words, given the variability associated with the process, queues are not always processed at equal speed. Furthermore, once users commit to a particular queue they are obligated to stay there even if they have to wait longer. It is assumed in this model that the rate of arrivals into each queue is not state dependent; that a new arrival is equally likely to join any queue regardless of relative length.

2.2.1 Limitations of the Model

This case represents the worst-case scenario and therefore produces the more conservative results. As explained above the fact that this model does not account for users selecting the shortest-queue server and no queue “jumping” is allowed differs from the normal behavior at toll plazas where vehicles attempt to select the shortest-queue lane and 'jockeying' from one queue to another is in some cases allowed.

2.3 Simulation of Queuing Using Traffic Simulation Software

No matter which mathematical formulation one selects, the analytical models depend on significant assumptions on the distributions of arrival and processing rates and on the type of dispatching disciplines. And they do not reflect the traffic dynamics of the approach to and departure from the toll plaza. To overcome these limitations, HNTB has developed applications of traffic simulation software to the toll plaza environment using VISSIM.

VISSIM^{iv} is a traffic simulation model currently gaining increasing acceptance in the transportation arena both in the U.S. and in the rest of the world. The model consists of two primary components that work in parallel:

- the traffic flow model
- the signal control model^v.

The simulation is microscopic (single vehicle modeling) and stochastic with fixed time-slices (one second intervals). The result of the simulation is an online animation of the traffic flow and offline reports of several traffic operation measures. Although links are used in the simulator, VISSIM does not have a traditional node-link structure. The lack of nodes provides the user with the flexibility to control traffic operations (e.g., yield conditions) and vehicle paths within complex road geometries.

Two models form the basis of the traffic flow model in VISSIM.

- the car-following model regulates how vehicles behave longitudinally
- the lane-change model, which uses a rule-based algorithm to govern lateral movements.

Generally, in the road, vehicles follow each other in an oscillating process. As a faster vehicle approaches a slower vehicle on a single lane, it has to decelerate. The time associated with this

reaction depends on the speed difference, distance and on driver's behavior. On multi-lane roadways, vehicles continuously check whether they can improve their current condition (traveling speed) by changing lanes. If so, they check the possibility of finding acceptable gaps on adjacent lanes. If available, the vehicle changes lanes and continues ahead.

2.3.1 Application of VISSIM to Toll Plazas

As mentioned before, toll plazas are not necessarily the most common applications of VISSIM, which was developed for limited access roadway and surface street networks. However, the model includes many capabilities that allow developing toll plaza simulations. Some of the particular capabilities of the model which are relevant in these applications are described below.

- **Dynamic assignment of vehicle paths**: the model includes a macro language that can be used to interact with vehicles during the simulation. This feature is particularly useful in toll plazas when vehicles need to be assigned to a particular tollbooth based on the current status of the system. This is accomplished by using in-road detectors on the network to detect the presence of vehicles approaching the plaza. Each second of the simulation, the program will check the status of detectors and assign the “most convenient” path to detected vehicles. The program keeps track of every vehicle assigned to each service lane, and of those who left the plaza. Additionally, every time a vehicle is assigned to a particular lane, the program will calculate a “perceived waiting time” based on the average service rate and the number of vehicle in that queue. In this manner, every second of the simulation the program will determine the status of the system and identify which lane is the most convenient for the next vehicle based on total vehicles in each queue and the estimated remaining time until the vehicle can be served. With this information, vehicles may be routed to the appropriate lane.
- **Priority Rules**: As explained before, the rules of the road or yielding conditions on the approach to and the exit from the plaza are somewhat different from other roadway segments. Because this is a lower-speed environment where more lane changes than usual take place, there is more awareness of drivers, thus more severe weaving is accepted as normal. The model allows coding priority rules that govern how vehicles will yield to each other. These can be related to either a time gap or a minimum distance between vehicles. This feature allows for a great flexibility in defining vehicle interactions.
- **Service Time Distribution**: this feature allows specifying either an empirical or a normal distribution of the processing time. This is especially important when toll transaction time data are available and thus a more realistic representation of the processing performance can be simulated. Unlike the analytical models, the service (transaction) times do not need to be Poisson distributed. Service time distributions can be associated with different vehicle classes (e.g., those paying with different payment methods – electronic toll collection, exact change or token, and cash for the toll plaza, and pay-on-foot, cashier, credit in/out, or electronic toll collection for parking exit) consequently allowing for differential processing rates.

- **Speed Reduction Zones**: differential desired speed distribution can be specified by mode, on a particular travel lane, therefore allowing for the simulation of toll lanes in which vehicles are not required to stop (e.g., electronic toll collection—ETC—lanes).
- **Driver Behavior**: several parameters in the model can be adjusted in order to modify how drivers interact and behave in the plaza. More aggressive or more cooperative behavior can be modeled to simulate real-life conditions.

Additionally, several MOEs can be obtained from the model that are relevant to the analysis of toll plazas, including:

- Average and maximum queue length by lane
- Average wait time by vehicle mode, by lane
- Average processing time
- Total throughput volume
- List of all vehicle delays measured over simulated period
- Total system delay
- Density and average speed on the approach and exit roadways

Note that these go beyond the four basic outputs of the mathematical models. Since the simulation model keeps track of every single vehicle in the system, a comprehensive list of statistics per vehicle is possible, allowing for very detailed analysis of the operation.

3. COMPARISON OF THE MODELS

Before applying the tools to a real case, they were used in the analysis of a series of generic plazas in order to compare the results and evaluate the validity and range of applicability of each model. For the application of the simulation model to this generic case, a configuration of the plaza was chosen so that there was no queue limitation. The purpose was to eliminate the influence of weaving, deficit of queuing area at the plaza, and other factors, and obtain a meaningful comparison of the tools under “ideal conditions”. The simulation model was run several times with different random seeds in order to capture traffic variability and different surge conditions. The same processing rate (360 vehicles per hour per lane) was used in all cases.

3.1 Single-queue comparison

The three models were tested under a single queue configuration. Table 1 shows results for this generic single-lane plaza. Because this is a single queue case, it is alleged that all models should yield similar results. This is in fact true for the analytical models that share the same formulation when number of servers equals one. However, the simulation results differ from the analytical models yielding less optimistic values for both queue length and wait times.

Arrival Rate (veh/hr)	M/M/1		VISSIM Simulation Results			
	Queue (veh)	Delay (sec/veh)	Queue (veh)	Delay (sec/veh)	Throughput	
					(veh/hr)	%
120	0.2	4.8	0	8.4	120	100.0%
240	1.3	19.8	14	260.4	219	91.3%
300	4.2	49.8	26	400.5	212	70.7%
350	34	349.8	82	1013.6	215	61.4%
355	70	709.8	83	1023.7	218	61.4%
359	358	3589.8	89	1042.1	222	61.8%
400	na	na	114.7	1190.6	220	55.0%

Table 1: Model Comparison for a Single-server Case

Figures 3 and 4 depict these results. As shown in these figures, the analytical tool provides more optimistic results for arrival rates below 97% of the maximum server capacity (v/c ratio < 0.97)

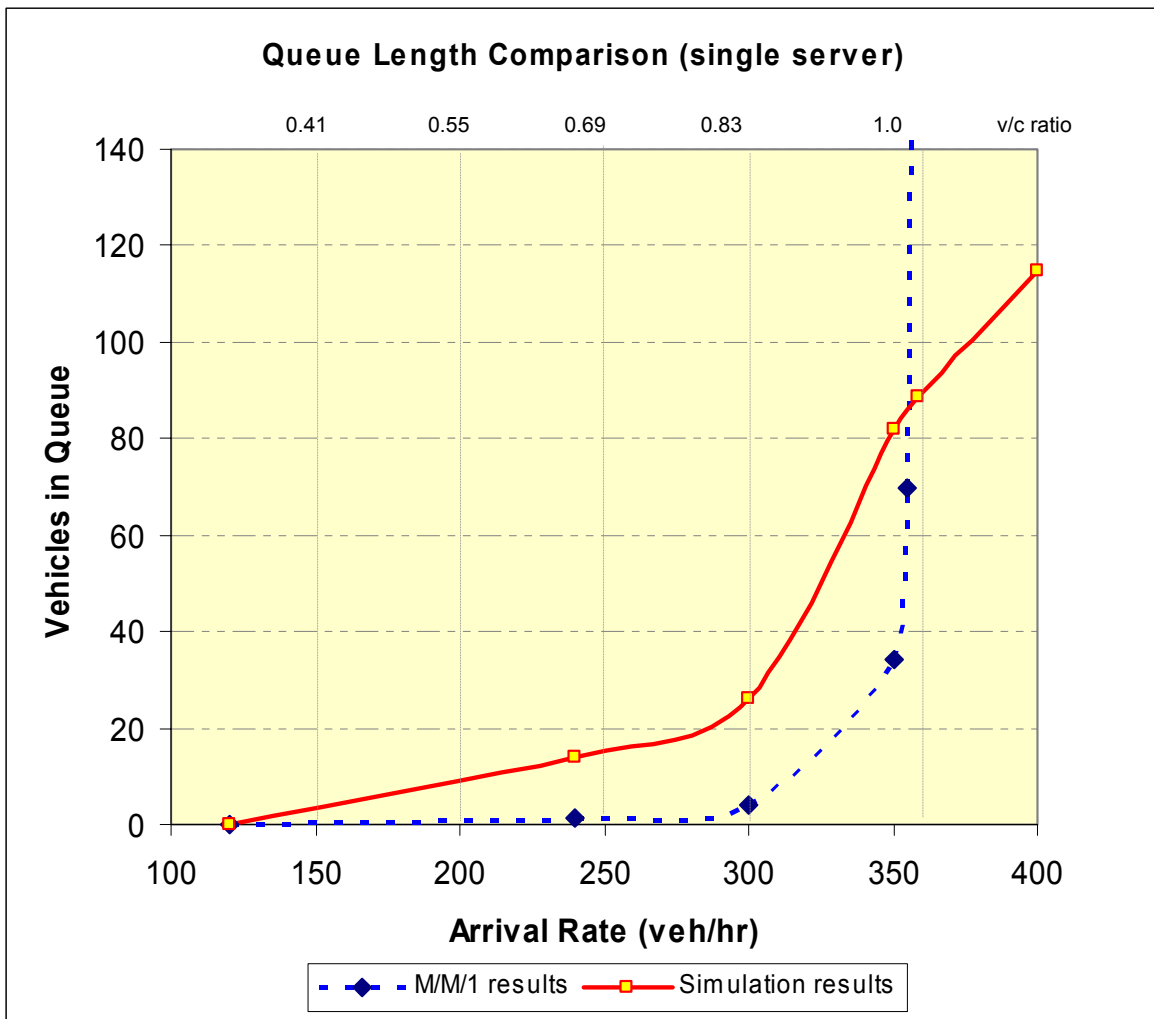


Figure 3: Comparison of Queue Length for the Single Server Case

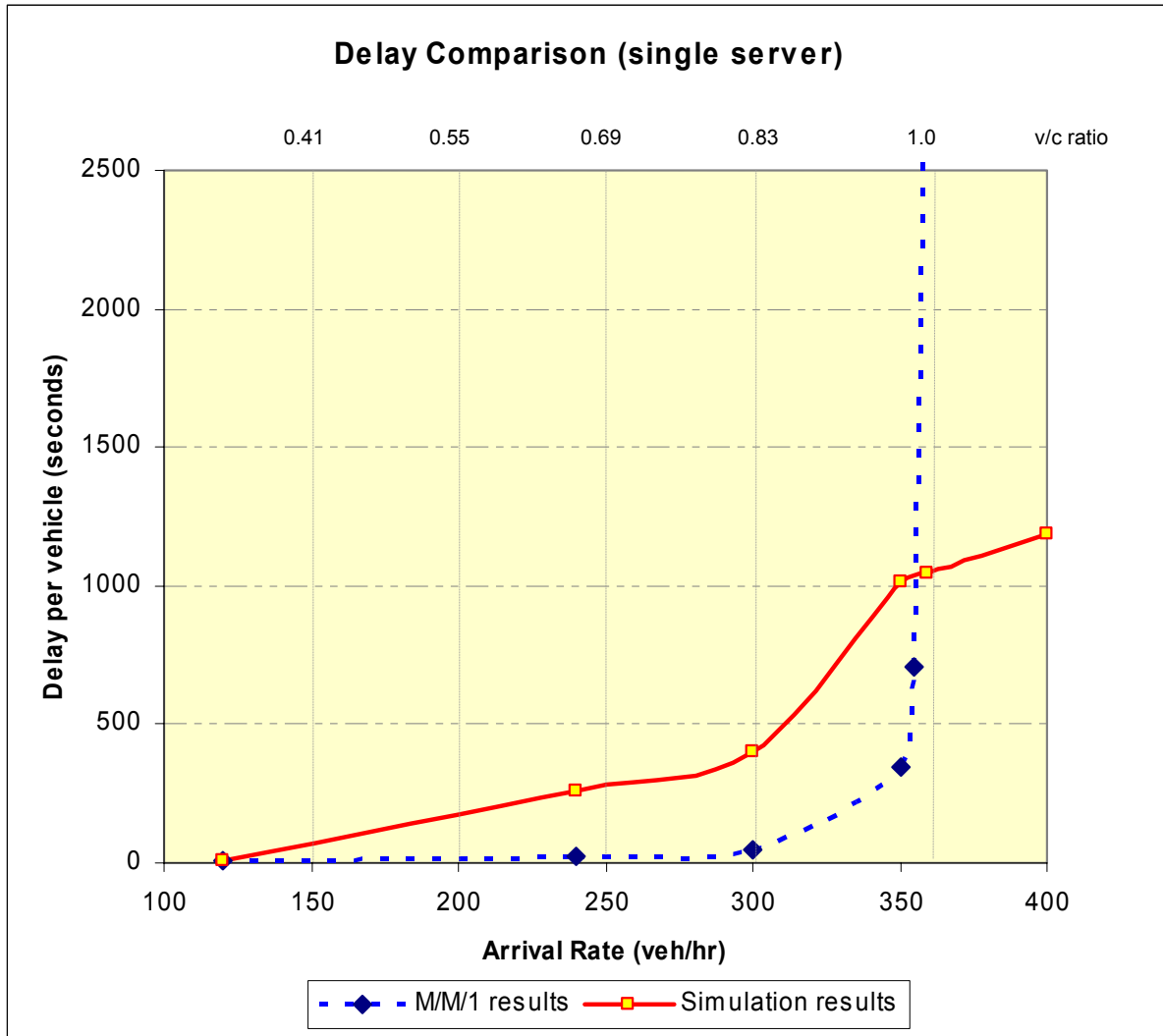


Figure 4: Comparison of Delay in Queue for the Single Server Case

One of the critical differences between the analytical and simulation models is the latter's ability to explicitly recognize the limits on throughput which result when queues build up. A review of this case shows that the maximum throughput reached by the simulation model (roughly 220 vehicles per hour) is less than the expected maximum capacity of the system at an average processing rate of 10 seconds per vehicle (360 vehicles per hour). Detail information from the simulation model showed that while the actual period of time vehicles were being processed averaged 10 seconds, the total time involved in the transaction, including deceleration to come to a stop condition and acceleration to leave the tollbooth, was greater than that. Total transaction time averaged 14.4 seconds per vehicle, which would yield a nominal capacity of roughly 250 vehicles per hour, very similar to the maximum throughput figure.

The reader may conclude that the analytical model could be adjusted to account for this bit of reality simply by re-running the model with the adjusted service rate. However, the results would show that the asymptotic point of the curves of queues and delays (where both go quickly to infinity) would simply shift to the left to the lower value where arrival rate equals service rate

(see Figures 7 and 8, and explanation below).

3.2 Multiple Queue Comparison

A similar procedure was followed to compare the analytical models with the simulation model under a multi-queue system. Table 2, Figure 5, and Figure 6 present the results for a generic two-lane plaza.

Arrival Rate (veh/hr)	M/M/s		2 x M/M/1		VISSIM Simulation Results			
	Queue (veh)	Delay (sec/veh)	Queue (veh)	Delay (sec/veh)	Queue (veh)	Delay (sec/veh)	Throughput	
							(veh/hr)	%
360	0.3	3.6	0.5	10.2	0.5	4.1	353	98.1%
450	0.8	6.6	1.0	16.8	7.0	124.6	430	95.6%
600	1.9	22.8	4.2	49.8	57.7	623.1	439	73.2%
650	4	43.8	8.4	93.0	70.2	939.0	439	67.5%
700	16.8	172.8	34.0	349.8	84.2	1038.8	439	62.7%
719	357.5	3592.8	358.0	3589.8	91.0	1070.0	439	61.1%
750	na	na	na	na	99.8	1115.0	439	58.5%

Table 2: Model Comparison for a Two-server Case

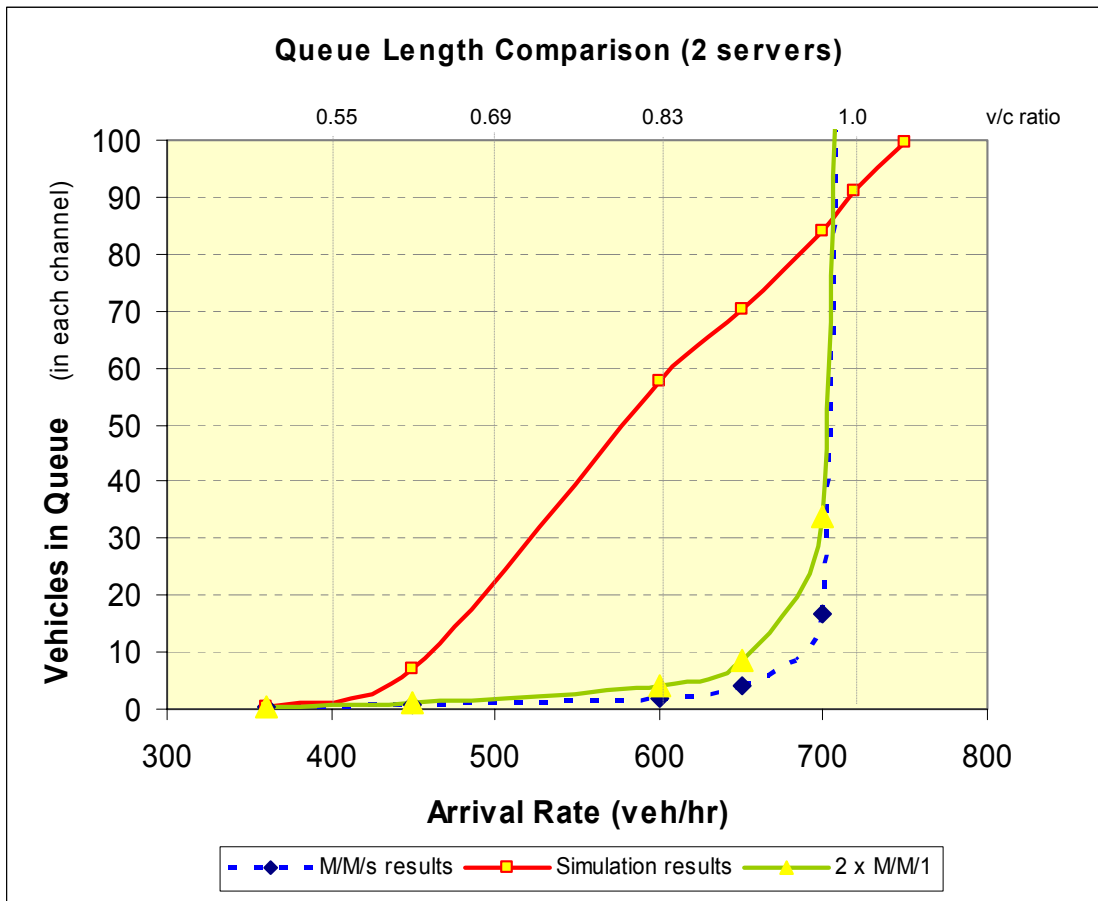


Figure 5: Comparison of Queue Length for a 2-server Case

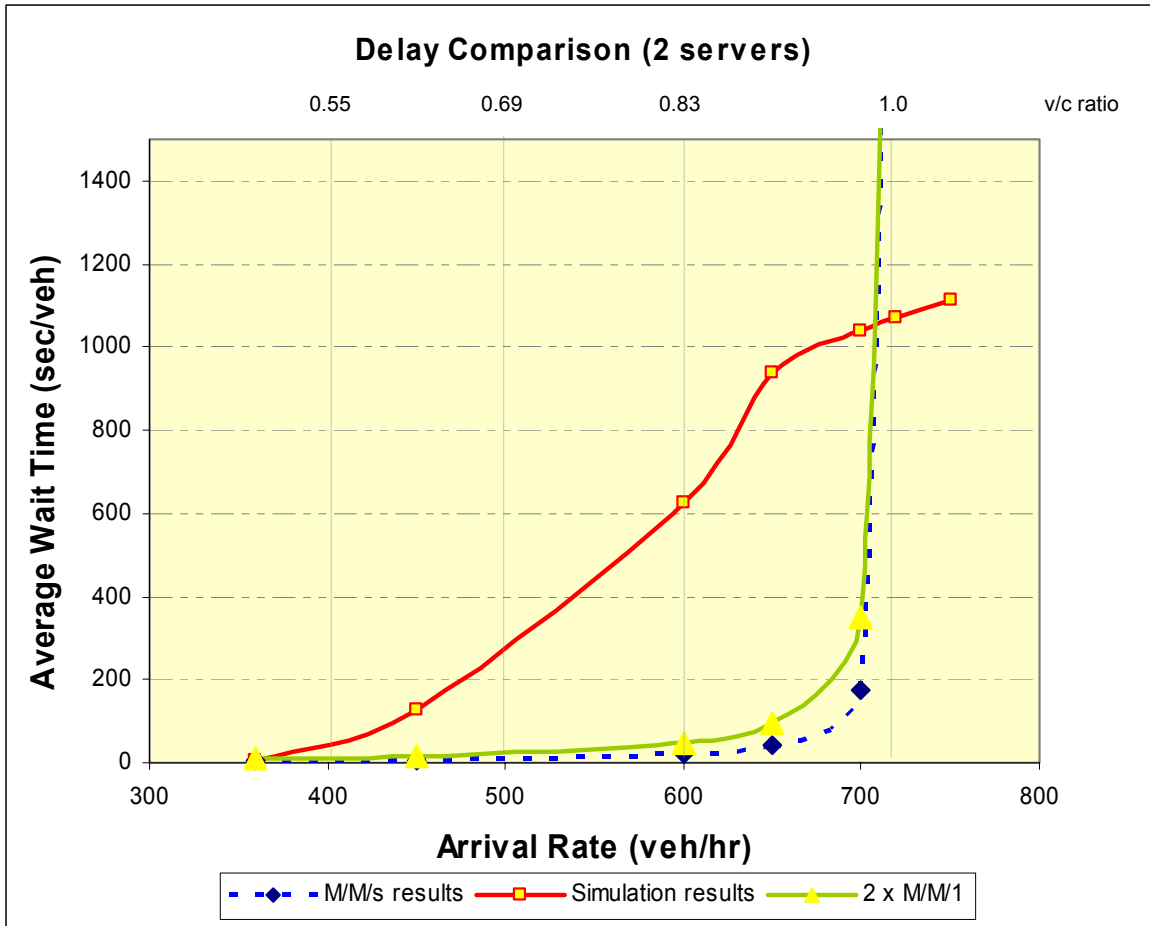


Figure 6: Comparison of Delay in Queue for a 2-server Case

The analysis of these values shows that all models differ in their results. The M/M/s results differ slightly from those for the multiple M/M/1, yielding more optimistic results. Both analytical methods greatly differ from the simulation results, with the MOEs from simulation significantly more conservative for v/c ratios less than 0.97.

Similarly to the single-server case, a review of the maximum throughput volumes processed by the simulation model demonstrates that the actual transaction times are greater than the defined processing or service times. Again, this is due to additional vehicle delays associated with approaching and leaving the service point (cashier booth, etc.), which in the end affects the overall capacity of the system.

When an increased processing rate (that accounts for actual processing time plus additional delays at the booth) is applied to the analytical models, the results show a smaller difference with simulation results for v/c ratios less than 0.9. Figure 7 shows the results of M/M/s and multiple M/M/1 when an average processing rate of 14.4 seconds is applied instead of the 10 seconds. The M/M/s model produced more optimistic results in terms of delay per vehicle while the multiple single-queue model produced values of delay above the simulation results. As showed in previous cases, when v/c ratios are greater than 0.97 neither analytical tool produce comparable results with the simulation model.

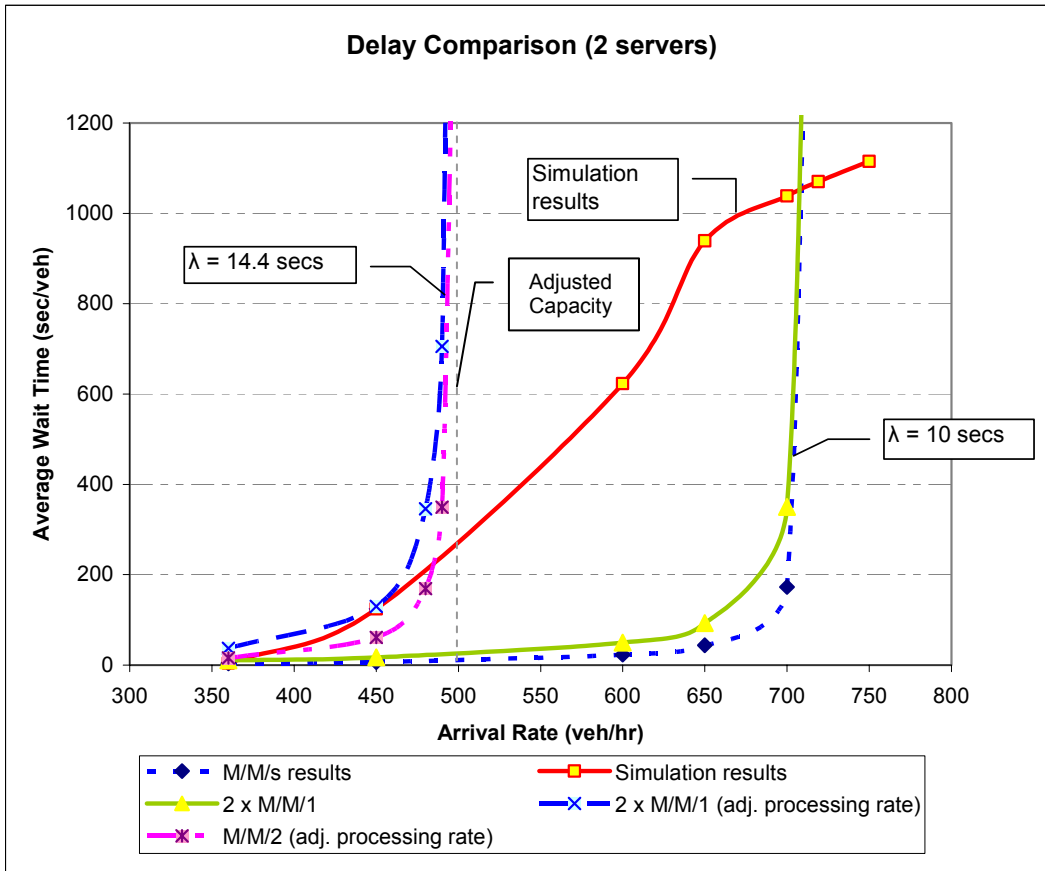


Figure 7: Wait Time Comparison for a Two-server Case with Adjusted Processing Rate

Similar comments can be drawn from the analysis of cases with more than two lanes. Figure 8 presents the results of a similar analysis for a plaza with six tollbooths.

4. VALIDATION OF THE SIMULATION MODEL

HNTB has applied both M/M/s and multiple M/M/1 queuing models along with simulation to several airport parking entry and exit plazas for both planning and evaluation purposes. One such case is reviewed in this paper. The parking exit plaza in the garage serving Terminals B and C at the Ronald Reagan Washington National Airport was used for the validation of the VISSIM model. This exit plaza serves both hourly and daily parking, is located internal to the garage, and is fed by several ramps from the five levels of the structure. As depicted in Figure 9, the existing configuration of the B/C parking garage exit plaza was coded into the simulation model. Data collected during the peak hour of a typical day of the week were used as inputs for arrival rates, service times, and lane use in the model.

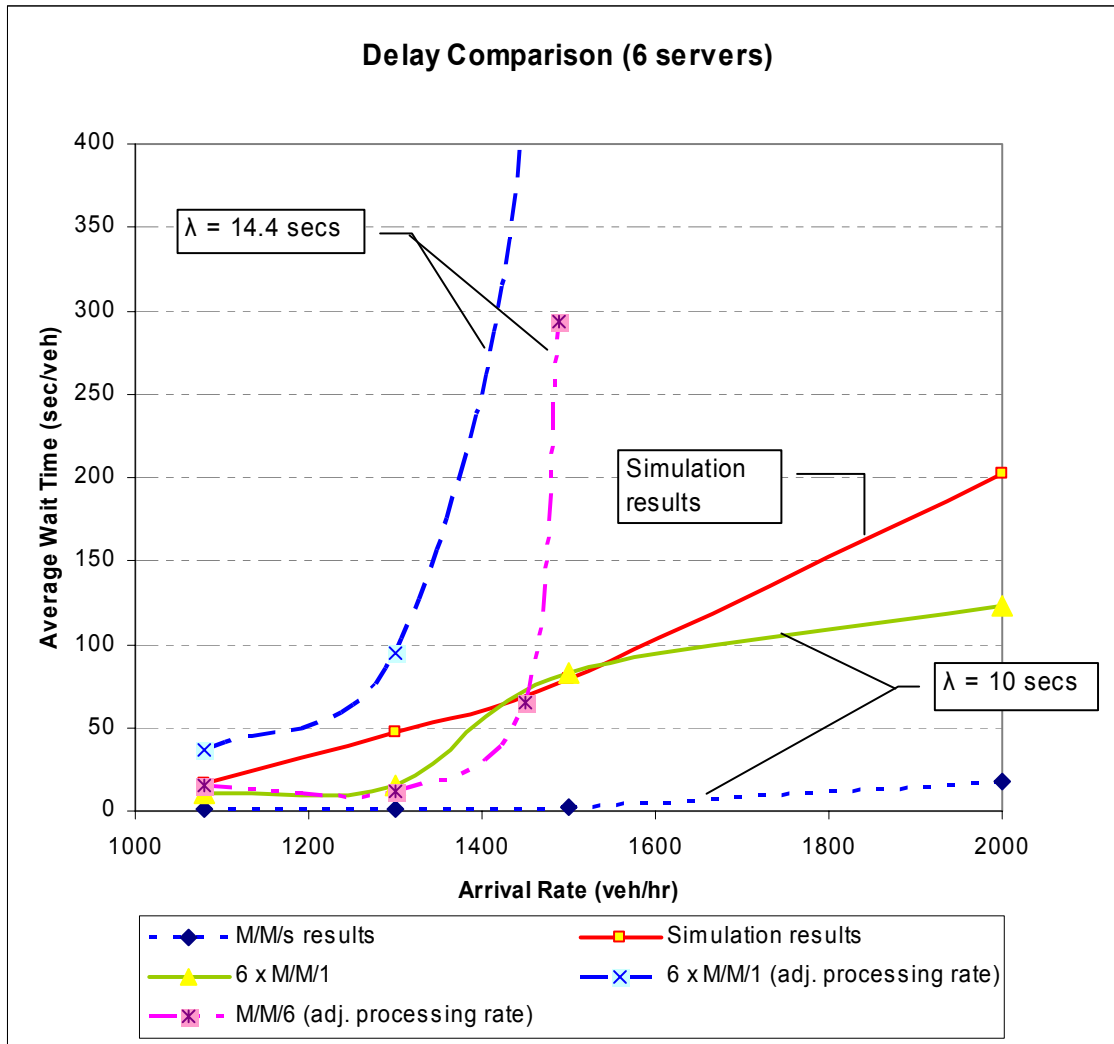


Figure 8. Wait Time Comparison with Six Servers

The analysis in this case focused on comparing the actual operation of the plaza with the simulation results. Particular emphasis was placed on defining the existing service time distribution. Figure 10 shows the empirical distribution of transaction times obtained on the field. A similar distribution was also coded into VISSIM as shown on Figure 11. It is important to note that while the service time averaged approximately 50 seconds per vehicle (a service rate of 72 vehicles per hour), there was a moderate percent of vehicles being stopped at the booth for more than two minute, and even in some cases, for more than 5 minutes. The upper tail of the transaction time cumulative distribution has a significant impact on the overall performance of the system since these few vehicles create a marked reduction in the capacity of the plaza.

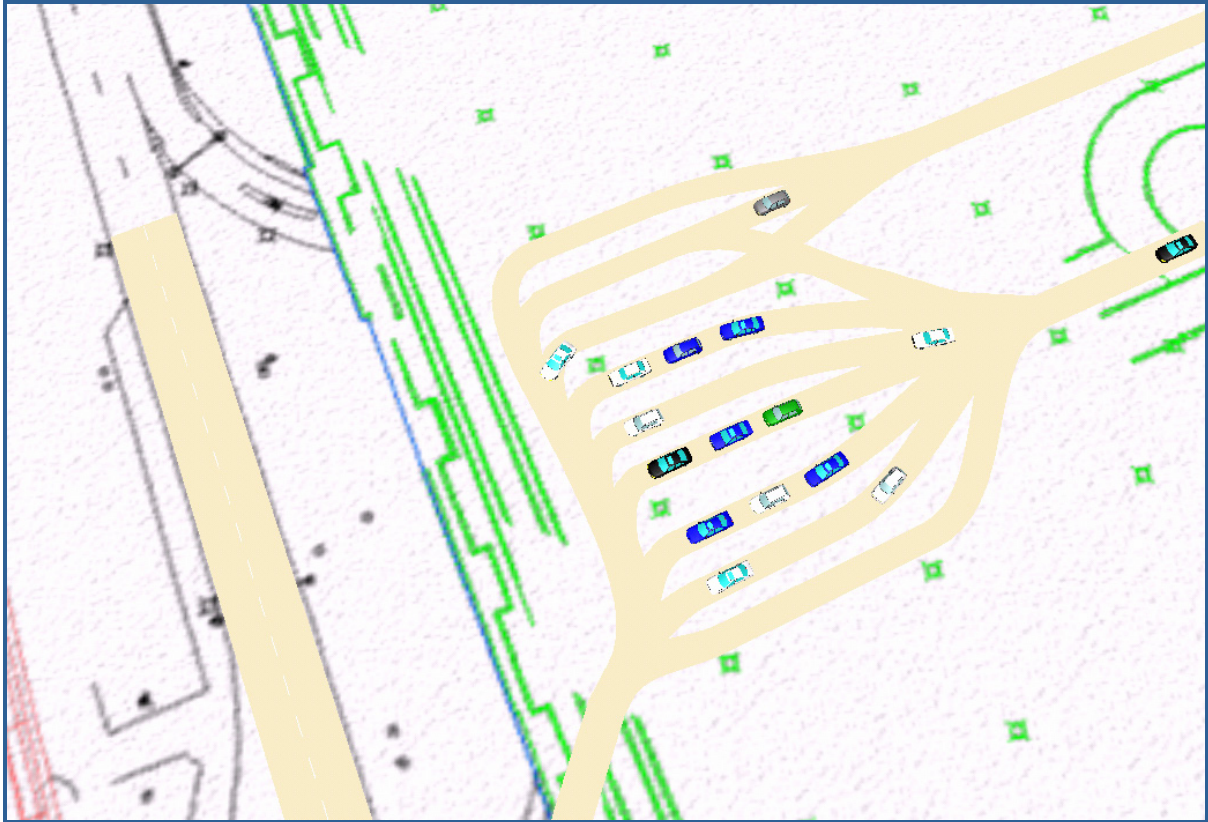


Figure 9: VISSIM Simulation of the B/C Parking Garage Exit Plaza

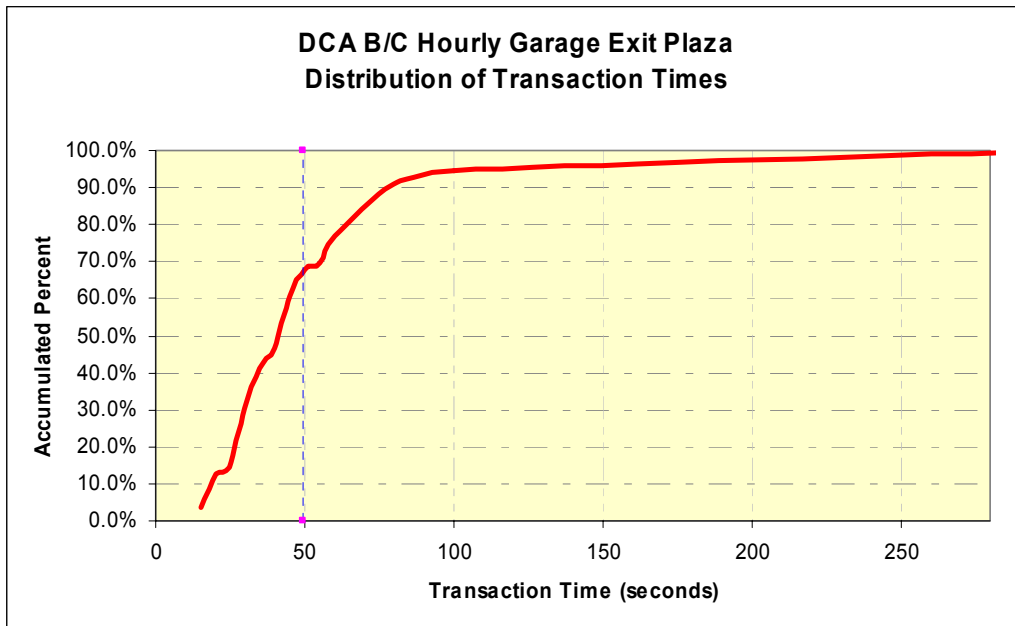


Figure 10: Collected Service Time Distribution

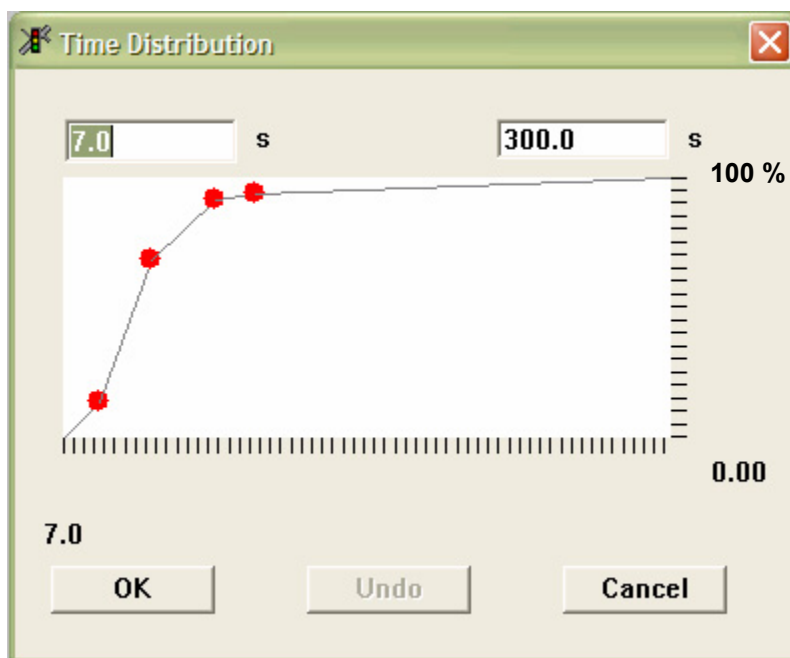


Figure 11: Service Time Distribution Coded in VISSIM

The analysis of the collected data showed that while there were nine lanes available at the plaza, only eight were in use at that time. Additionally, there was significant lane-use imbalance due to the particular plaza configuration. Roughly 80 percent of vehicles approaching the exit lanes had a very short distance (< 150 feet) in which to select the shortest-queue lane. Visibility and maneuverability are restricted due to the presence of columns. Therefore, the central lanes of the plaza were selected more frequently. This operating constraint was coded into the simulation model by assigning levels of preference to each lane.

As in previous cases, the model was run several times with different random seeds and MOEs were collected during an hour of simulation. The comparison of results focused on the following MOEs:

- Throughput volume
- Average and maximum queues on each lane
- Average wait time in queue.

Figures 12 through 15 show the results of comparing collected versus simulated data. Figure 12 illustrates this comparison for throughput volumes by 15-minute time slices. It can be concluded from these results that the simulation model is processing the correct number of vehicles throughout the entire period.

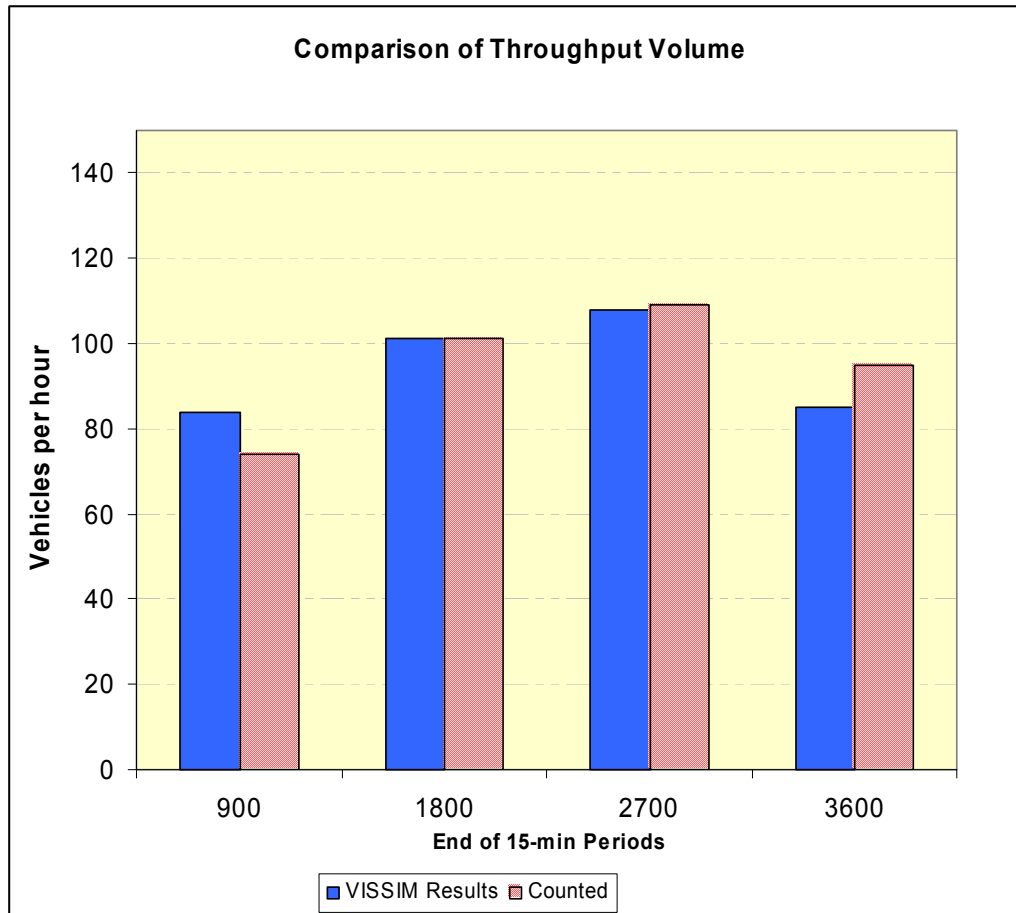


Figure 12: Throughput Volume Comparison

Similarly, Figure 13 presents the comparison for average queue length. The reader can note the imbalance of lane use at the plaza where, because of its particular configuration, central lanes are more heavily used than others. The results in Figure 13 demonstrate that the simulation model does an adequate job in replicating this particular driver behavior. The figure also shows a comparison with the expected queue lengths obtained by using the multi-server and multiple single-server queue models. Two separate systems comprising six lanes and two lanes respectively were analyzed to account for the particular operation of this plaza. Again, results confirmed that multiple single servers produce better estimates of queue length. However, the goodness of these estimates is dependent on the total number of servers of the system. Indeed, the higher the number of servers, the more conservative the estimate. Multi-server analysis produced significantly more optimistic results in all the cases.

Figure 14 compares maximum queue length by lane observed at the plaza with simulation results. As in the previous cases, the simulation model adequately replicated real-life queuing conditions and lane use of the plaza.

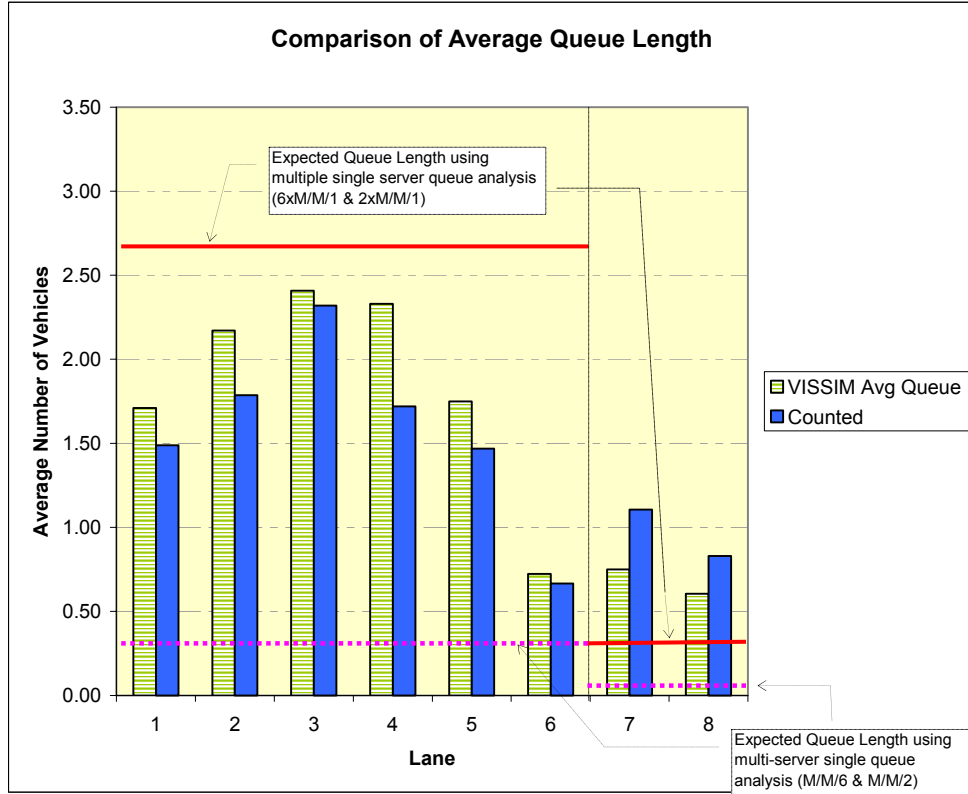


Figure 13: Queue Length Comparison between Counts and Simulation Results

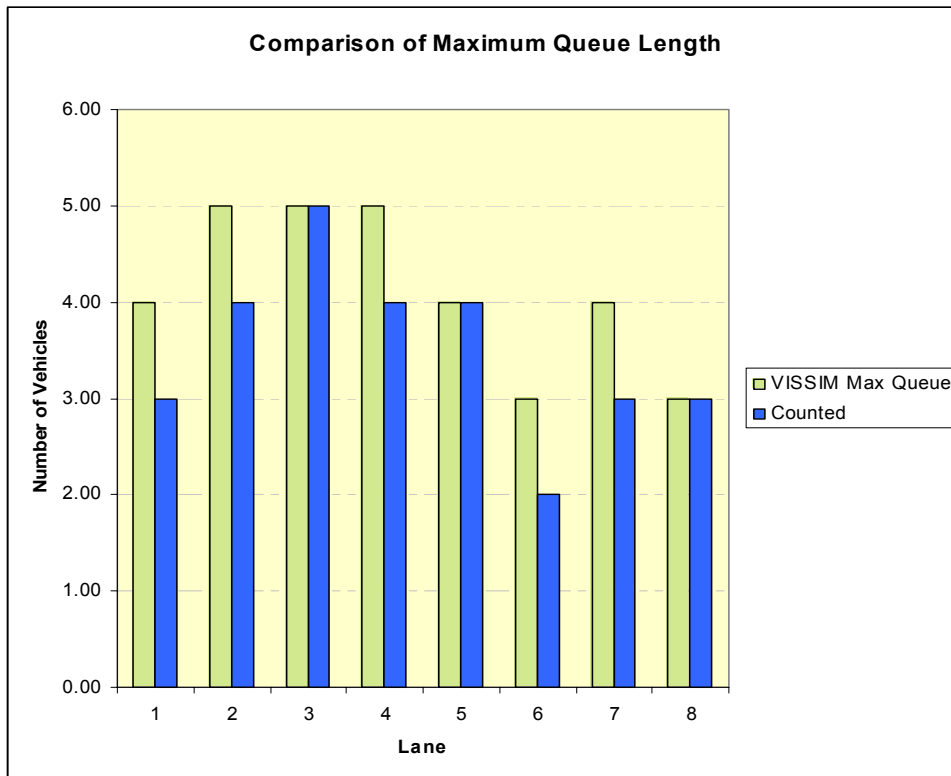


Figure 14: Comparison of Maximum Queue Length

Figure 15 summarizes the comparison between average time in queue per vehicle for each service lane of the plaza. As shown by this figure, simulation results reasonably matched the values measured on the field both in terms magnitude and distribution. On the other hand, the analytical queuing models are not as accurate estimators of the time in queue. The multi-server, single queue model produced significantly more optimistic values for the expected wait time than observed. In contrast, the multiple single-server model produced values for this variable slightly above the maximum values measured in the field. This can be considered acceptable for planning purposes.

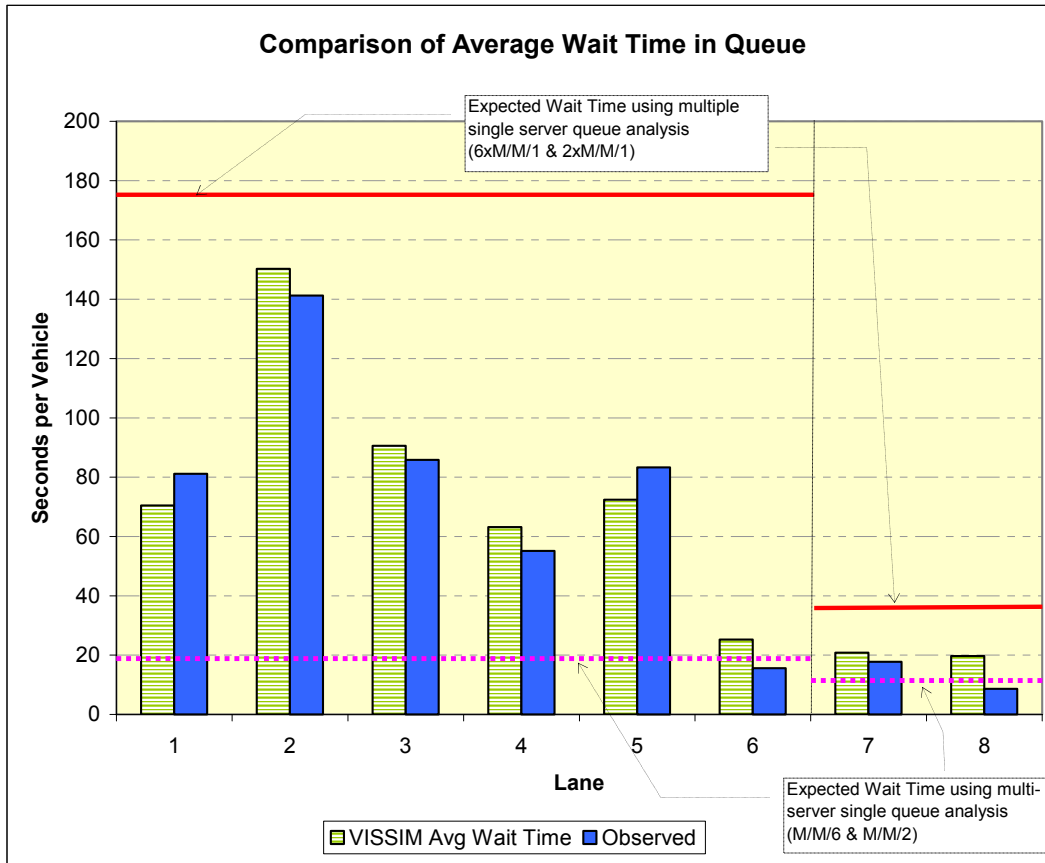


Figure 15: Comparison of Average Wait Time

5. FINDINGS AND CONCLUSIONS

5.1 Applicability of Multi-server Analytical Models

The analysis and comparison presented in this paper leads the authors to the conclusion that good judgment must be employed when using the analytical models, as their results may differ significantly from real-life plaza operation. It is important to understand the limitations and range of applicability of these tools. On the one hand, multi-server single queue models tend to

produce optimistic results, which could lead to erroneous estimates when obtaining toll plaza requirements or evaluating its operation. On the other hand, multiple single-server models may produce estimations closer to values observed in actual operation, however these estimations can vary from optimistic when the number of servers tend to one, to very pessimistic for eight or more servers.

In all cases, use of the analytical models should be avoided when trying to analyze toll plaza operation under high levels of demand ($0.90 < v/c \text{ ratios} < 1.0$), given their asymptotic behavior within this range.

5.2 Applicability of the VISSIM Simulation Model

The VISSIM simulation model is well adapted to simulate the operation of a toll plaza. Comparison with observed values of queue lengths, throughput volumes, and wait times shows that reasonably good estimations can be produced with this model. Available capabilities in VISSIM allow for an accurate representation of the toll plaza configuration including the ability to consider lane-use imbalance, and other particular user behaviors. The array of MOEs that can be extracted from the model allows for in-depth analyses of the plaza operation. Additionally, both traffic arrival and service time distributions can be coded into the model to match real-life operation. Calibration and validation of the model are critical to assure that the appropriate driver behavior and particular constraints of the plaza under analysis have been captured in the model.

Simulation models take more time to develop and to produce results than using the analytical queuing models here described. HNTB has considerably reduced the up-front development time needed for this kind of application, being in the order of a few days depending on how complex the plaza configuration may be.

In the roadway environment (see next section), an advantage of simulation is the fact that the toll plaza model can be integrated with the rest of the roadway network it serves. By linking a model of roadway operation with the toll plaza model, the latter will inherit the operational characteristics of the upstream traffic. This allows for a more realistic and systematic representation of the overall transportation system.

5.3 Applicability to Roadway Toll Plazas

The applications of the VISSIM simulation to queuing phenomena at fee plazas described in this paper were all parking exit plazas at airports. While the authors have described these as “toll” plazas, it is useful to consider the differences between the parking exit plaza environment and that found at toll plazas on toll roadways, bridges, and tunnels. Among the differences which are most relevant are:

- **scale of the demand:** the largest airport parking garages in the United States are less than 20,000 spaces, and serve both hourly as well as daily parking demand. Typical durations for the daily parking are 2 – 4 days, resulting in low turnover per space and small peak hour demands. Hourly spaces typically turn over at most once per hour, and

average durations may be up to 2 – 3 hours. Thus while these spaces generate considerably more peak demand on the exit plaza than daily parking, the total peak hour demand from a very large garage may be only in the range of 1000 – 2000 vehicles. By contrast, 2,000 vehicles per hour **per lane** is the peak hour demand typical of a busy freeway; a six-lane freeway would therefore generate perhaps 6,000 vehicles in the peak hour, peak direction. The scale of the plazas therefore tend to be larger and the system busier, making it more difficult for drivers both to estimate the preferred lane, as well as to maneuver to it.

- **speed of the approaching traffic:** Parking garages have approach speeds of less than 15 mph, and typically in the 5 – 10 mph range. On a freeway in rush hour near (but not at capacity), such approach speeds will likely be 45 – 60 mph. Considerable signing is provided to slow traffic down on approaches to toll barriers, and heavy traffic and long queues help to achieve this. But the result is still higher approach speeds than for a parking plaza. The implication of higher speeds is in the ability to convey the information to the driver about the condition of the system, including which lanes are available, which ones have longer queues, which are moving faster, and which accept which types of payment. Simply put, speed complicates the scale issue, magnifying the difficulty of the driver having the information he needs to make the wisest choice, and increasing his difficulty in maneuvering to achieve that choice.
- **duration of peak times:** Peak times at airport parking garage exits are somewhat dependent on flight schedules, which dampen the maximum ability of the terminal system to send passengers to the garage and out the exit. Thus peaks at garages tend to come and go quickly as surges within the peak hour. On a freeway in a congested metropolitan area, on the other hand, “rush hour” conditions may last for several hours of uninterrupted high demand. When surges within this long rush hour happen repeatedly over several hours, the toll plaza may not have the ability to catch up and empty the queues, which therefore build back up faster when the next surge arrives.

While it is clear therefore that the parking and the roadway (toll road, bridge, tunnel) cases differ, the simulation model should still be applicable. Each of these characteristics can be represented in a simulation approach using VISSIM. Given its traffic operations simulation origins, it can reflect the interaction between vehicles and the driver’s behavior in responding to the information presented to the driver by the toll plaza system. What is essential is that the analyst validate the model first using data gathered at the toll plaza in question before modifying the model to analyze future conditions which may represent changes in plaza configuration, operation, or both.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Planning Department of the Metropolitan Washington Airports Authority for providing parking data and necessary support during field work. Additionally, the authors would like to acknowledge Mr. Matthew Houston and Mr. Ross Burton for their help during data collection.

Author's Information:

Gustavo A. Ceballos
Senior Transportation Planning Engineer
HNTB Corporation
2900 South Quincy St. Suite 200
Arlington, VA 22206-2265
Phone: 703-824-5100 x2629, Fax: 703-671-6210
Email: gceballos@hntb.com

Owen P. Curtis
Associate Vice President
HNTB Corporation
2900 South Quincy St. Suite 200
Arlington, VA 22206-2265
Phone: 703-824-5100 x2629, Fax: 703-671-6210
Email: ocurtis@hntb.com

Notes:

ⁱ A stochastic process has the Markov property if the probability distribution of future states of the process depends only upon the current state, and are conditionally independent of the past states given the present state. A process with the Markov property is usually called a Markov process.

ⁱⁱ Kendall's Notation was introduced in 1953. It describes the three main characteristics of queuing systems: (a/b/c), where "a" represents the arrival distribution, "b" represents the service time distribution, and "c" represents the number of parallel servers in the system.

^{iv} VISSIM is a commercially available traffic simulation package developed by PTV AG, Karlsruhe, Germany, and distributed in the United States by PTV America, Inc., Corvallis, OR.

^v Due to the inapplicability of signal control in the toll or exit plaza environment, this model is not discussed here.