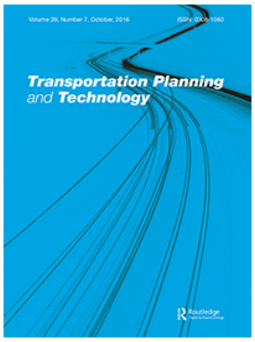

Routledge
Taylor & Francis Group

Transportation Planning and Technology

ISSN: 0308-1060 (Print) 1029-0354 (Online) Journal homepage: http://www.tandfonline.com/loi/gtpt20

Analysing queueing at toll plazas using a coupled, multiple-queue, queueing system model: application to toll plaza design

Partha Chakroborty, Rahul Gill & Pranamesh Chakraborty

TRANSPORTATION PLANNING AND TECHNOLOGY, 2016
VOL. 39, NO. 7, 675–692
http://dx.doi.org/10.1080/03081060.2016.1204090

Analysing queueing at toll plazas using a coupled, multiple-queue, queueing system model: application to toll plaza design

Partha Chakroborty, Rahul Gill and Pranamesh Chakraborty

Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh, India

ABSTRACT

A vehicle approaching a toll plaza observes the queues at each of the available toll-lanes before choosing which to join. This choice process, the arrival process of vehicles and the service characteristics of the toll-booths, affect the queues and delay the drivers. In this paper, queueing at a toll plaza is modelled as a multiple-queue queueing system where the arrival process to a queue (toll-lane) is dependent on the state of all the queues. In the past, such systems have been modelled mathematically only for two queues and are not applicable for toll plazas with three or more toll-lanes. The proposed model determines the steady-state probability density function (pdf) for the queues at large toll plazas. This study is used to determine the number of toll-lanes or the length of the upstream queueing area required to achieve certain user-specified levels-of-service. Expected delay and maximum queue length are used as level-of-service measures. Indicative design charts are also provided.

ARTICLE HISTORY

Received 25 May 2015
Accepted 12 May 2016

KEYWORDS

Toll plaza design; toll plaza queueing model; coupled parallel queues; steady-state probability density function; level-of-service

1. Introduction

Toll plazas are required so that the revenue for the maintenance and improvement of road infrastructure can be generated, at least in part, from those who use it. Yet by their existence they form impediments to the smooth flow of traffic; it is this unavoidable inefficiency that must be minimized. There are many examples where improperly designed toll plazas cause severe inconvenience to travellers and reduce the mobility benefits that are supposed to accrue from an expressway. The purpose of this paper is to develop a realistic model for queueing at toll plazas so that such facilities can be designed efficiently given the flow on the approach expressway and service characteristics of the toll-lanes (or toll-booths).

Over the years various computer simulation and analytical (queueing theory) models have been attempted. The computer simulation models include those by Redding and Junga (1992), Al-Deek, Mohamed, and Radwan (2000), Correa, Metzner, and Nino (2004), and Russo, Harb, and Radwan (2010). These models, available at various levels of sophistication, although useful for studying toll plazas, cannot be considered as a replacement for analytical or queueing theory models that provide steady-state probability density functions for the queues that form at a toll plaza.

CONTACT Partha Chakroborty partha@iitk.ac.in

Edie (1954) was one of the first to attempt the use of queueing theory in analysing the delay at a toll plaza. He used results from multiple server single-queue queueing systems, like M/M/S and M/D/S (Taylor and Karlin 1984), where there is one queue that leads to all the *S* servers and the user at the top of the queue joins the next available server. Clearly, this is not how vehicles queue at a toll plaza. Sometime later, Haight (1958) published his analysis of a queueing system applicable for a two toll-lane toll plaza. In the analysis, he considers two queues in parallel with the queuers joining the shortest queue. In this paper, he correctly avoids looking at the queueing at a toll plaza as a multiple server, single-queue, queueing system; instead, he models it as a parallel, multiple-queue, queueing system.

Schwartz (1972) also points out how queueing at toll plazas are of the kind where there are multiple queues and arriving vehicles choose one of them after evaluating the state of all the queues. Schwartz (1972) further indicates that this choice, in the general case, can be predicted only probabilistically and as a function of the queue lengths of each of the queues (or toll-lanes) that the vehicle could have joined. He also presents analyses for some special cases (with less than or equal to three servers or toll-lanes) where choice of a queue takes place deterministically according to some rules. Conolly (1984) compares how the system will fare had the queueing at a toll plaza not been the kind it is, but if it truly were a multiple server, single-queue, queueing system. He, like Haight (1958), restricts the discussion to only a two-server (two toll-lane) system.

As the previous discussion shows, it has been long realized that one cannot use M/M/S, M/D/S or similar such queueing analysis to analyse the queueing process at a toll plaza. However, and somewhat surprisingly, the single-queue queueing model continues to be used even in recent works (cf. Kim 2009). One of the reasons for this, although not justifiable, could be the fact that the multiple, parallel-queue models of Haight (1958) and Schwartz (1972) are extremely difficult to solve. Schwartz (1972), for example, mentions that the procedure described by him is 'fantastically complicated'. Blanc (1987), while discussing similar queueing systems, albeit in a different setting, also mentions that 'queueing systems with more than one waiting line are in general hard to analyze'.

This paper formulates the queueing process at a toll plaza as a multiple server, multiple parallel-queue queueing system (as opposed to a multiple server, single-queue queueing system like M/M/S) with approaching drivers choosing a queue or toll-lane. The paper applies a power series algorithm (PSA) described by Blanc (1987) to determine the steady-state, state probabilities of the queueing system (that is, the probabilities of queue lengths on the different toll-lanes). A closed-form analytical solution to a special case of the proposed queueing system is also used to see how effectively the PSA determines the state probabilities. The results show that the PSA-based method is accurate.

As an application of the proposed model, the obtained state probabilities are used to develop indicative tables and figures on the required minimum number of toll-lanes and minimum length of the queueing area (upstream of the toll plaza) for different arrival and service rates. In order to gain further confidence in the proposed model and its solution technique, VISSIM, a popular, commercially available microscopic simulation tool, is also used to simulate the flow at toll plazas. From the simulations of flow under various combinations of arrival and service rates, the minimum number of toll-lanes required for each of these combinations is also determined. These are then compared with those obtained from the proposed model for the same arrival and service rates. The comparisons show that the proposed model realistically describes the queueing process. It may be pointed out that the tables provided here are only to indicate how the proposed model can be used to design essential elements of a toll plaza and are not meant to be exhaustive.

The paper is divided into four further sections. The next section describes the problem more precisely and presents a mathematical formulation. The third section describes how the PSA is used to determine the state probabilities. The fourth section presents results to (i) validate the proposed model and its PSA-based solution procedure and (ii) indicate how the state probabilities can be used to determine minimum number of toll-lanes or the minimum length of upstream queueing areas required to provide some user-specified level of service. The final section concludes the paper by highlighting the contributions and the shortcomings of the present work. This section also points out the areas that require further attention.

2. Problem statement and formulation

Figure 1 shows a typical toll plaza. The essential features of the toll plaza are the number and type of toll-lanes (or toll-booths) and the length of the queue area (or alternatively the length of the widened section) immediately upstream of the toll-booths. This length acts as a storage space for the stopped vehicles waiting in queues to pay the toll. Vehicles of different types arrive at the plaza. Each vehicle evaluates the queues in the various toll-lanes it can join, chooses one toll-lane, and joins it. Sometimes, a vehicle also changes from its original toll-lane to one of the adjacent toll-lanes.

There is ample evidence that drivers go through a choice process before selecting a toll-lane; one can refer to Gulewicz and Danko (1995), Mudigonda, Bartin, and Ozbay (2009), and Dubedi et al. (2012) for more discussions on the choice process at a toll plaza. The fact that vehicles choose a toll-lane (or queue) after evaluating the queue lengths of all the toll-lanes implies that the arrival process to a queue (or toll-lane) is dependent on the condition (length) of the other queues at the toll plaza. This choice process creates a dependence or coupling between the parallel queues of the system. This is true even if

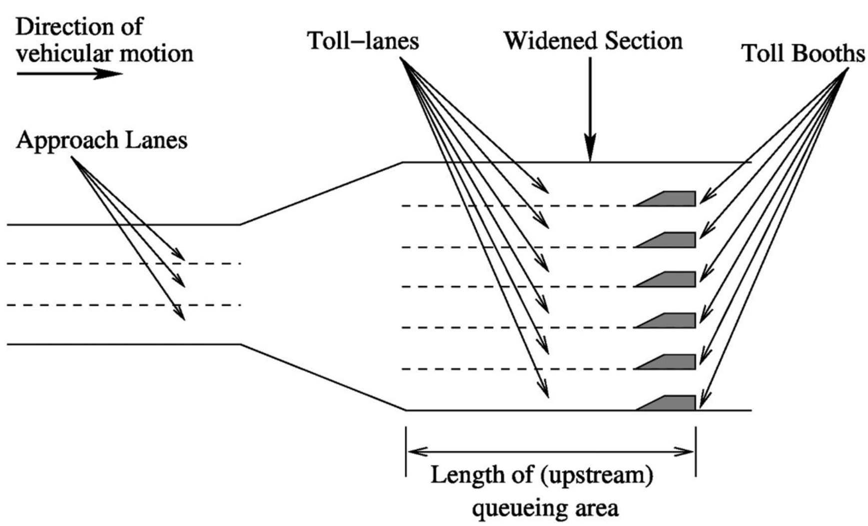

Figure 1. Schematic diagram of a typical toll plaza.

one ignores the reneging of vehicles from one queue (toll-lane) to join another. Hence, in this paper, the queueing system at the toll plaza is referred to as a coupled, multiple-queue, queueing system (CMQ2S). It is felt that such a terminology describes the queueing system at a toll plaza more closely than the 'queueing system with more than one waiting line' terminology of Blanc (1987) or 'lane selecting queueing models' terminology of Schwartz (1972) , or 'parallel queues' terminology of Haight (1958).

In the rest of the section, a mathematical formulation for the CMQ2S at a toll plaza is presented. The description and formulation of the problem presented here frequently uses the phrase 'state of the system'. The state of the queueing system at a given time is described as a vector of the queue lengths at the different toll-lanes (of the toll plaza) at that time. For example, if a toll plaza has four toll-lanes and at a given time the queue length on the first toll-lane is, say, 2, on the second toll-lane 7, on the third toll-lane 4, and on the fourth toll-lane 5, then the state of the system at that time is (2, 7, 4, 5). Furthermore, the term queue length of a toll-lane, as used here, includes the vehicle being served at the toll-booth.

Before proceeding further, the assumptions used in the formulation and the notation are presented. The assumptions are:

(1) Vehicles arrive according to the Poisson distribution,
(2) Service times at the toll-booths are distributed exponentially,
(3) An arriving vehicle chooses a toll-lane or queue depending on the state of the system (i.e. depending on the queue lengths at all the toll-lanes),
(4) The choice process can be assumed to be deterministic like, 'join the shortest queue' or stochastic wherein only a probability that a vehicle will join a queue is available,
(5) Once a vehicle joins a toll-lane or queue it cannot switch to another toll-lane or queue, and
(6) The type of arriving vehicle has no impact on the queueing system.

The notation is as follows:

T	Total number of toll-lanes or toll-booths (servers) at the toll plaza
N_i	A random variable denoting the number of vehicles (including the vehicle paying the toll) in the queue on toll-lane i
n_i	The values N_i can take (these are non-negative integers)
$\bar{n}$	The vector $(n_1, n_2, n_3, \ldots, n_T)$ which describes the state of the system; note that the state of the system at any time is described only in terms of the queue lengths at each of the toll-lanes
$\lvert\bar{n}\rvert$	A scalar indicating the sum of individual elements of the vector $\bar{n}$, i.e. $\lvert\bar{n}\rvert = n_1 + n_2 + n_3 + \ldots + n_T$, the total number of vehicles in the system
l_i	Length of queue on toll-lane i
λ	The rate at which vehicles arrive at the toll plaza
λ_i	The rate at which vehicles join toll-lane i (or arrive at toll-lane i)
μ	The rate at which vehicles depart from the toll plaza
μ_i	The rate at which vehicles depart from toll-lane i
ρ	The ratio of λ to μ
ϕ_i	The ratio of μ_i to μ
$P(\rho, \bar{n})$	Probability of finding the system in state $\bar{n}$, i.e. probability $[N_i = n_i;\ i = 1 \ \ldots T]$; it is also a function of ρ
$\pi_i(\bar{n})$	The probability that an arriving vehicle chooses toll-lane (or queue) i when the system is in state $\bar{n}$
$\Gamma(a)$	A function which returns a value of 1 if argument a is true, else returns a value of zero
$\bar{\eta}_i$	A vector $(\eta_1, \eta_2, \ldots, \eta_T)$ with $\eta_i = 1$ and $\eta_j = 0\ \forall j \neq i$
p_i	Proportion of vehicle type i in the traffic stream
L_i	Length of vehicle type i

It can be said that when the CMQ2S reaches a steady state (i.e. the state probabilities do not vary over time), the following equation must hold:

$$\left\{\sum_{i=1}^{T}\pi_i(\bar{n})\lambda + \sum_{i=1}^{T}\mu_i\Gamma(n_i > 0)\right\}P(\rho, \bar{n}) = \left\{\sum_{i=1}^{T}\mu_i P(\rho, \bar{n} + \bar{\eta}_i)\right\} + \left\{\sum_{i=1}^{T}\pi_i(\bar{n} - \bar{\eta}_i)\lambda\Gamma(n_i > 0)P(\rho, \bar{n} - \bar{\eta}_i)\right\} \tag{1}$$

This equation indicates that in the steady (stationary) state the probability of leaving a particular state (as expressed by the LHS of Equation (1)) must be equal to the probability of entering that state (as expressed by the RHS of Equation (1)). The LHS indicates that the probability of leaving a particular state is equal to the probability of either an arrival or a departure occurring while the system is in that state. Note that a departure can occur from a particular queue if and only if that queue has at least one vehicle. The RHS is derived based on the fact that one can enter a particular state either through a departure from the queue (or toll-lane) or through an arrival to the queue (or toll-lane) that made the present state different from the state the system is entering into. (Two points should be noted: (i) no queue can ever be in a state with negative number of vehicles and (ii) arrival and departure processes are assumed to be Poisson.)

Dividing throughout by μ and substituting $\phi_i = \mu_i/\mu$, Equation (1) can be re-written in terms of ρ (where $\rho = \lambda/\mu$) as:

$$\left\{\rho\sum_{i=1}^{T}\pi_i(\bar{n}) + \sum_{i=1}^{T}\phi_i\Gamma(n_i > 0)\right\}P(\rho, \bar{n}) = \left\{\sum_{i=1}^{T}\phi_i P(\rho, \bar{n} + \bar{\eta}_i)\right\} + \left\{\rho\sum_{i=1}^{T}\pi_i(\bar{n} - \bar{\eta}_i)\Gamma(n_i > 0)P(\rho, \bar{n} - \bar{\eta}_i)\right\} \tag{2}$$

For a given ρ, equations such as Equation (2) and the relation indicating that the sum of all $P(\rho, \bar{n})$ is unity, can be solved to yield the state probabilities, $P(\rho, \bar{n})$. However, solving this set of equations is difficult. Blanc (1987) presents a PSA method based on the ideas of Keane, Hooghiemstra and van de Ree (referred to in Blanc (1987) as private communications) that can be used to obtain the state probabilities for the present queueing system represented by Equation (2). Blanc's procedure is a numerical scheme based on power series expansions of state probabilities as a function of ρ. This method has been chosen here because 'experience has taught that the algorithm is more powerful than algorithms based on truncation of the state space, and that it provides more accurate results' (Blanc 1992). The next section describes how the state probabilities are determined in this paper using the PSA. The discussion is brief and provided mainly for purposes of completeness. The interested reader may refer to Blanc (1987, 1992) and Hooghiemstra, Keane, and Van De Ree (1988) for detailed understanding of the PSA method for obtaining the state probabilities.

Before leaving this section two issues related to Equation (2) need to be discussed. The first concerns the determination of ϕ_i, and the second the determination of $\pi_i(\bar{n})$.

Although the formulation presented implies that the value of ϕ_i can be different for every i (in fact, it could easily depend on the state of the system also), in the numerical examples presented in the rest of the paper it has been assumed that all toll-booths are of the same kind and hence have the same ϕ_i. Thus, $\mu = T\mu_i$ and $\phi_i = 1/T$.

As mentioned earlier, $\pi_i(\bar{n})$ is the probability that an arriving vehicle chooses toll-lane i (or the queue for toll-booth i) when the state of the system is $\bar{n}$. In the absence of a proper model on how people choose a toll-lane, analysts generally assume that drivers choose the toll-lane with the shortest queue. That is, if S is the set of toll-lanes (or queues) with the shortest queue length and $|S|$ is the cardinality of the set S, then under this assumption

$$\pi_i(\bar{n}) = \begin{cases} 1/|s| & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

The above expression for $\pi_i(\bar{n})$ can be used in the proposed formulation to determine the state probabilities. However, such a quasi-deterministic (note if $|S| = 1$ then $\pi_i(\bar{n}) =$ 1 or 0) way of looking at $\pi_i(\bar{n})$ is not desirable. Ideally, $\pi_i(\bar{n})$ should be determined through a proper model of the toll-lane choice process. A good way of modelling the toll-lane choice process, wherein a driver chooses from a set of mutually exclusive and collectively exhaustive set of alternatives (or toll-lanes), is through discrete choice analysis (DCA).

In the DCA, a decision-maker (in this case a driver) is modelled as one who evaluates each alternative (in this case the queue at a toll-lane) on a utility scale and then chooses the one that provides the maximum utility. However, the utility is typically understood as a random quantity (an idea first introduced by Thurstone 1927) and the probability that an alternative has the greatest utility is assumed to be the probability that the alternative will be chosen from among all the alternatives. The utility is assumed to be a sum of two parts: the systematic part, typically represented as a function of certain measurable explanatory variables, and a random part. Under specific assumptions about the nature of the randomness in the utility one gets the Logit model. The interested reader may refer to Ben-Akiva and Lerman (1985) for a detailed exposition on this topic.

In this paper, the recommended strategy is to determine $\pi_i(\bar{n})$ through a Logit model:

$$\pi_i(\bar{n}) = \frac{e^{V_i(\bar{n})}}{\sum\limits_{\forall i} e^{V_i(\bar{n})}}. \tag{4}$$

Here, $V_i(\bar{n})$ is the systematic part of the (scaled) utility function used to determine the utility of toll-lane i and is estimated using actual observations on the toll-lane choice process. Extensive data collected from three toll plazas on Indian expressways (more than 700 choices are observed) are used to estimate $V_i(\bar{n})$. The $V_i(\bar{n})$ obtained from the data and used in this paper is presented in the section on numerical examples.

3. Determination of state probabilities, $P(\rho, \bar{n})$

As mentioned earlier, the PSA method developed and explained in Blanc (1987) is used here to determine $P(\rho, \bar{n})$ that satisfies equations like Equation (2) with $\phi_i = 1/T$ and

$\pi_i(\bar{n})$given by Equation (4) ($\pi_i(\bar{n})$ given by Equation (3) can also be used). The method relies on expressing $P(\rho, \bar{n})$ in Equation (2) as the following power series in terms of ρ:

$$P(\rho, \bar{n}) = \rho^{|\bar{n}|} \sum_{k=0}^{\infty} \rho^k c(k, \bar{n}), \tag{5}$$

where $c(k, \bar{n})$ are the coefficients of the power series. Using the expression obtained by substituting $P(\rho, \bar{n})$from Equation (5) into Equation (2) and the law of total probability, one can recursively calculate the coefficients, $c(k, \bar{n})$. The process is described in detail in Blanc (1987) together with a bilinear mapping approach which helps in the convergence of the PSA for higher values of ρ. In this, ρ is replaced by the following expression in θ:

$$\rho = \frac{\theta}{1 + G - G\theta}, \tag{6}$$

where G is a non-negative constant whose value has to be chosen through trial and error (in a later paper Blanc (1993) provides some ideas as to how G can be chosen). These ideas have been implemented in this paper through a computer program in order to calculate the coefficients with some degree of accuracy. Blanc (1993) states that the biggest problem with this method is not so much computation time but the amount of storage space required for the coefficients. However, it may be noted that even though storage space remains the primary concern, computation time is not negligible and increases sharply with the number of servers (or toll-lanes).

The next section presents results from various numerical examples or cases.

4. Results

This section has two primary purposes: first, to establish the efficacy of the proposed model by comparing its results with simulation results and results from analytical methods (that work under some particular assumptions); and, second, to indicate how the results from the proposed method can be used to design toll plazas.

The results from these cases are divided into two broad categories and presented in the following two subsections. The first subsection presents results from studies that are aimed at establishing that the proposed model and its PSA-based solution methodology help evaluate the steady-state probabilities reasonably accurately. The second subsection presents results from various cases that are different in terms of the arrival rates, service rates, etc. The purpose here is to show how the proposed model can be used to determine the required number of toll-lanes or required size of the queueing area for different input conditions. The designs obtained from the proposed method are also compared with those obtained from VISSIM, a commercially available micro-simulation model. The results show a close match.

4.1. Tests to validate the proposed methodology

As with any numerical scheme, it is important to see how accurately the PSA-based method used here determines the state probabilities. In order to see whether the method as well as its implementation is working satisfactorily, two tests are done. The first is analytical in nature, whereas in the second VISSIM is used. The results of the second test are included in the next subsection in order to improve readability.

In the first test case, it is assumed that $\pi_i(\bar{n}) = 1/T$. Given that arrival and departure processes are assumed to be Poisson, this assumption on $\pi_i(\bar{n})$ implies that the queueing system, in effect, operates as T independent M/M/1 queues with an arrival rate of λ/T to each of them. Hence, the state probabilities of any of the queues can be obtained analytically and these could then be compared with the corresponding state probabilities obtained using the PSA-based method.

It may be noted that the probability of finding a queue (say on toll-lane i) in a particular state (say, n_i) for a given value of ρ, $P(\rho, n_i)$, can be obtained as a marginal distribution from the state probabilities, $P(\rho, \bar{n})$ calculated using the PSA-based method. That is, noting that

$$P(\rho, \bar{n}) \equiv P(\rho, n_1, \cdots, n_i, \cdots, n_{\mathrm{T}}), \tag{7}$$

one can obtain the marginal probabilities, $P(\rho, n_i)$ as

$$P(\rho, n_i) = \sum_{n_T=0}^{\infty} \cdots \sum_{n_{i+1}=0}^{\infty} \sum_{n_{i-1}=0}^{\infty} \cdots \sum_{n_1=0}^{\infty} P(\rho, n_1, \cdots, n_i, \cdots, n_{\mathrm{T}}). \tag{8}$$

Furthermore, since under the assumptions of $\pi_i(\bar{n}) = 1/T$ and Poisson arrival and departure processes the CMQ2S operates as T independent M/M/1 queues, one should expect that all the marginal distributions obtained using $P(\rho, \bar{n})$ are identical. That is, $P(\rho, n_1)$, $P(\rho, n_2)$, … $P(\rho, n_{\mathrm{T}})$ should all be identical; or in other words, for a given ρ

$$P(\rho, n_1 = a) = P(\rho, n_2 = a) = \cdots = P(\rho, n_{\mathrm{T}} = a) \quad \forall a. \tag{9}$$

It is observed that marginal probabilities from the PSA method obtained using Equation (8) satisfy the condition stated in Equation (9). That is, the probability distributions for all the queues are identical. Table 1 presents a comparison of the state probabilities calculated using the M/M/1 analysis (arrival rate of λ/T) with the

Table 1. Comparison of state probabilities.

	ρ values							
	0.4		0.6		0.7		0.8	
n_i	M/M/1	PSA	M/M/1	PSA	M/M/1	PSA	M/M/1	PSA
0	0.6000	0.6000	0.4000	0.4000	0.3000	0.3000	0.1996	0.2000
1	0.2400	0.2400	0.2400	0.2400	0.2100	0.2100	0.1598	0.1600
2	0.0960	0.0960	0.1440	0.1440	0.1470	0.1470	0.1279	0.1280
3	0.0384	0.0384	0.0864	0.0864	0.1029	0.1029	0.1023	0.1024
4	0.0154	0.0154	0.0518	0.0518	0.0720	0.0720	0.0819	0.0819
5	0.0061	0.0061	0.0311	0.0311	0.0504	0.0504	0.0656	0.0655
6	0.0025	0.0025	0.0187	0.0187	0.0353	0.0353	0.0525	0.0524
7	0.0010	0.0010	0.0112	0.0112	0.0247	0.0247	0.0420	0.0419
8	0.0004	0.0004	0.0067	0.0067	0.0173	0.0173	0.0337	0.0336
9	0.0002	0.0002	0.0040	0.0040	0.0121	0.0121	0.0270	0.0268
10	0.0001	0.0001	0.0024	0.0024	0.0085	0.0085	0.0216	0.0215
11	0.0000	0.0000	0.0015	0.0015	0.0059	0.0059	0.0173	0.0172
12	0.0000	0.0000	0.0009	0.0009	0.0042	0.0042	0.0138	0.0137
13	0.0000	0.0000	0.0005	0.0005	0.0029	0.0029	0.0111	0.0110
14	0.0000	0.0000	0.0003	0.0003	0.0020	0.0020	0.0089	0.0088
15	0.0000	0.0000	0.0002	0.0002	0.0014	0.0014	0.0071	0.0070

marginal probabilities obtained using Equation (8). The table has only one column for different n_i values since the marginal probabilities for different n_i are identical. The values from the M/M/1 analysis are listed under the columns titled M/M/1 and the values from PSA approximation of CMQ2S are listed under the columns titled PSA. As seen from the table, the values are identical up to the fourth decimal place with slight differences arising for a ρ value of 0.8. At even higher values of ρ (although not shown here) these differences increase but are within acceptable limits. (Higher values of ρ are not included since, in general, systems are never designed for values of ρ higher than 0.8.).

4.2. Applications in toll plaza design and further validation

The primary design parameters at a toll plaza are: (i) the number of toll-lanes (or toll-booths) that should be provided, (ii) the relative distribution of manual versus automatic toll-lanes (this, however, has not been investigated here) and (iii) the length of the upstream queue area at the toll plaza (see also Figure 1) so that stopped vehicles do not spill over to the expressway lanes. The criterion on which these designs are based is (can be) one of the following: (i) the maximum queue length at the toll plaza (referred to as the maximum queue length criterion) should be below a user-defined threshold or (ii) the average waiting time faced by the users of the toll plaza (referred to as the waiting time criterion) should be below a user-defined threshold.

Since queue length is a stochastic quantity, while using the maximum queue length criterion, the design parameters should be chosen such that the probability of a queue length at the toll plaza exceeding the user-defined maximum is below a threshold. The average waiting time, on the other hand, is a deterministic quantity and can be used directly to obtain the design parameters. In this case, the design parameters should be chosen such that the average or expected waiting time of a vehicle at the toll plaza is below a user-defined threshold average waiting time.

As mentioned earlier, in this paper the required number of toll-lanes or required length of the upstream queueing area is calculated with $\pi_i(\bar{n})$ given by a logit model (see Equation (4)). The systematic part of the utility function for a given toll-lane, say i, is assumed to depend only on the queue length on that lane. That is, $V_i(\bar{n})$, the systematic utility derived by a driver from toll-lane, i is:

$$V_i(\bar{n}) = kn_i,$$

where k is a real constant. The value of k is estimated from more than 700 observations on toll-lane choice (by drivers) at three different toll plazas in India. The maximum likelihood estimate of k is -0.25.

Furthermore, the number of toll-lanes or the length of the upstream queueing area calculated here assumes that all the toll-lanes are of the same kind (even though the steady-state queueing equation of Equation (2) is generic and can handle different types of toll-lanes). This restriction became inevitable because the $V_i(\bar{n})$ is estimated from data available for only one kind of toll-lanes, namely manual.

In the following, the calculation procedures and the associated tables/figures for the number of toll-lanes and the length of upstream queueing area are presented.

4.2.1. Determination of number of toll-lanes using the maximum queue length criterion and further validation

In this section, the adequate number of toll-lanes (recall that all are of the same kind) at a toll plaza based on the maximum queue length criterion is calculated. It is proposed that the number of toll-lanes should be such that the probability of any of the queue lengths being greater than a user-defined 'acceptable queue length, Q' is less than some threshold, α.

Let the probability of any queue length being greater than Q, be Ω. This probability depends on λ, μ_i and T (since the state probabilities depend on ρ) and Q. Hence, Ω is written as $\Omega(\lambda, \mu_i, T, Q)$ and is given as:

$$\Omega(\lambda, \mu_i, T, Q) = \sum_{\forall \bar{n} \in N} P(\rho, \bar{n}), \tag{10}$$

where set N is defined as

$$N \equiv \{n_1, n_2, \cdots, n_T | n_1 > Q \text{ or } n_2 > Q \cdots \text{or } n_T > Q\},$$

and state probabilities, $P(\rho, \bar{n})$ are determined using the procedure described in the previous sections.

Thus, one has to determine the minimum value of T, say T^*, which satisfies $\Omega(\lambda, \mu_i, T, Q) \leq \alpha$. It may be noted that the reason for introducing λ and μ_i in place of ρ is that, from an engineering perspective, it is more natural to talk about arrival rates and service rates. The values of T^* for different values of λ, μ_i and Q and for $\alpha = 0.05$ are provided in Tables 2–6. Also note that, since all the toll-lanes are the same, all μ_is are also same. Another point that needs to be mentioned is that for the computation resources available at the disposal of the authors and given the computation-/memory-intensive nature of the PSA-based solution algorithm, the state probabilities could be accurately determined for up to a maximum of eight servers (toll-lanes). Hence, in many cases the table reads '≥ 9' indicating that results suggest that eight servers are not sufficient.

In order to validate the results further, an application on VISSIM is developed to simulate the queueing at the toll plazas. In this, only automobiles are considered and it is

Table 2. Minimum number of toll-lanes required (T^*) when using the maximum queue length criterion; $\mu_i = 250$ vph (per toll-lane).

	λ (vph)				
	250	500	750	1000	1250
Q	S, PSA	S, PSA	S, PSA	S, PSA	S, PSA
3	2, 3	4, 5	5, 8	7, ≥ 9	8, ≥ 9
4	2, 2	3, 4	5, 6	6, 8	7, ≥ 9
5	2, 2	3, 3	4, 5	6, 7	7, 8
6	2, 2	3, 3	4, 4	6, 6	6, 7
7	2, 2	3, 3	4, 4	5, 5	6, 6
8	2, 2	3, 3	4, 4	5, 5	6, 6
9	2, 2	3, 3	4, 4	5, 5	6, 6
10	2, 2	3, 3	4, 4	5, 5	6, 6
11	2, 2	3, 3	4, 4	5, 5	6, 6
12	2, 2	3, 3	4, 4	5, 5	6, 6
13	2, 2	3, 3	4, 4	5, 5	6, 6
14	2, 2	3, 3	4, 4	5, 5	6, 6
15	2, 2	3, 3	4, 4	5, 5	6, 6

Table 3. Minimum number of toll-lanes required (T^*) when using the maximum queue length criterion; $\mu_i = 500$ vph (per toll-lane).

	λ (vph)							
	250	500	750	1000	1250	1500	1750	2000
Q	S, PSA	S, PSA	S, PSA	S, PSA	S, PSA	S, PSA	S, PSA	S, PSA
3	1, 1	2, 3	3, 4	3, 5	4, 7	5, 8	5, ≥9	6, ≥ 9
4	1, 1	2, 2	2, 3	3, 4	4, 5	4, 6	5, 7	5, 8
5	1, 1	2, 2	2, 3	3, 3	4, 4	4, 5	5, 6	5, 7
6	1, 1	2, 2	2, 3	3, 3	3, 4	4, 4	5, 5	5, 6
7	1, 1	2, 2	2, 2	3, 3	3, 4	4, 4	5, 5	5, 5
8	1, 1	2, 2	2, 2	3, 3	3, 4	4, 4	5, 5	5, 5
9	1, 1	2, 2	2, 2	3, 3	3, 3	4, 4	4, 5	5, 5
10	1, 1	2, 2	2, 2	3, 3	3, 3	4, 4	4, 4	5, 5
11	1, 1	2, 2	2, 2	3, 3	3, 3	4, 4	4, 4	5, 5
12	1, 1	2, 2	2, 2	3, 3	3, 3	4, 4	4, 4	5, 5
13	1, 1	2, 2	2, 2	3, 3	3, 3	4, 4	4, 4	5, 5
14	1, 1	2, 2	2, 2	3, 3	3, 3	4, 4	4, 4	5, 5
15	1, 1	2, 2	2, 2	3, 3	3, 3	4, 4	4, 4	5, 5

assumed that they choose a toll-lane based on the probabilities given in Equation (4). The simulation results are used to determine the number of toll-lanes using the maximum queue length criterion in much the same way it is done for the PSA procedure. Tables 2 and 3 provide the values of T^* obtained using simulation along with the values obtained using the PSA-based method for CMQ2S (comparison for only two values of μ are presented to avoid repetitions and also because the comparisons presented allow one to infer that the values based on the proposed method match well with the simulated values, especially for realistic values of Q). In these two tables, each cell contains two values of T^* reported in the format *S*, *PSA*; the value under *S* is from the VISSIM-based simulation while that under PSA is from the PSA-based method for the CMQ2S.

Tables 2–6 show that T^* behaves as per expectations; it increases as λ increases and decreases as μ_i increases or Q increases. These results can be directly used to determine the minimum number of toll-lanes required at a toll plaza given the arrival rate, service rate and acceptable queue length values. For example, for a service rate of 500 vph (per toll-lane), arrival rate of 1500 vph and a Q value of 4, the minimum number of toll-lanes required, T^*, can be read from Tables 3 as 6.

Table 4. Minimum number of toll-lanes required (T^*) when using the maximum queue length criterion; $\mu_i = 750$ vph (per toll-lane).

	λ (vph)						
Q	500	750	1000	1500	2000	2500	3500
3	2	3	3	5	7	≥ 9	≥ 9
4	2	2	3	4	5	7	≥ 9
5	1	2	3	3	4	5	7
6	1	2	2	3	4	5	7
7	1	2	2	3	4	5	6
8	1	2	2	3	4	4	6
9	1	2	2	3	4	4	6
10	1	2	2	3	4	4	6
11	1	2	2	3	3	4	6
12	1	2	2	3	3	4	6
13	1	2	2	3	3	4	6
14	1	2	2	3	3	4	6
15	1	2	2	3	3	4	6

Table 5. Minimum number of toll-lanes required (T^*) when using the maximum queue length criterion; $\mu_i = 1000$ vph (per toll-lane).

	λ (vph)						
Q	500	750	1000	1500	2500	3500	4500
3	2	2	3	4	7	≥9	≥9
4	1	2	2	3	5	7	≥9
5	1	2	2	3	4	6	7
6	1	1	2	3	4	5	6
7	1	1	2	2	4	5	6
8	1	1	2	2	4	5	6
9	1	1	2	2	3	5	5
10	1	1	2	2	3	4	5
11	1	1	2	2	3	4	5
12	1	1	2	2	3	4	5
13	1	1	2	2	3	4	5
14	1	1	2	2	3	4	5
15	1	1	2	2	3	4	5

Table 6. Minimum number of toll-lanes required (T^*) when using the maximum queue length criterion; $\mu_i = 1250$ vph (per toll-lane).

	λ (vph)						
Q	500	750	1000	1500	2500	3500	4500
3	1	2	2	3	5	8	≥9
4	1	2	2	3	4	6	7
5	1	1	2	2	3	5	6
6	1	1	2	2	3	4	5
7	1	1	2	2	3	4	5
8	1	1	2	2	3	4	5
9	1	1	2	2	3	4	5
10	1	1	2	2	3	4	5
11	1	1	2	2	3	4	4
12	1	1	2	2	3	4	4
13	1	1	1	2	3	3	4
14	1	1	1	2	3	3	4
15	1	1	1	2	3	3	4

4.2.2. Determination of number of toll-lanes using the waiting time criterion

The number of toll-lanes at a toll plaza can also be determined as the minimum number of toll-lanes required to ensure that the expected waiting time of vehicles at the toll plaza is less than some user-defined threshold, W. The expected waiting time in the system (i.e. expected time between the time of leaving the toll plaza and the time of joining a queue/toll-lane at the toll plaza), $E[W]$, can be obtained as:

$$E[W] = \frac{E[|\bar{n}|]}{\lambda}, \tag{11}$$

where $E[|\bar{n}|]$ is the expected number of vehicles queued in the system (or the toll plaza) and can be calculated easily from $P(\rho, \bar{n})$. One can refer to any standard text book on stochastic processes, like Taylor and Karlin (1984), for a discussion on (i) how to obtain expected number in the system, $E[|\bar{n}|]$ from the state probabilities, and (ii) the relation presented as Equation (11). Of course, $E[W]$ is a function of (λ, μ_i, T) because

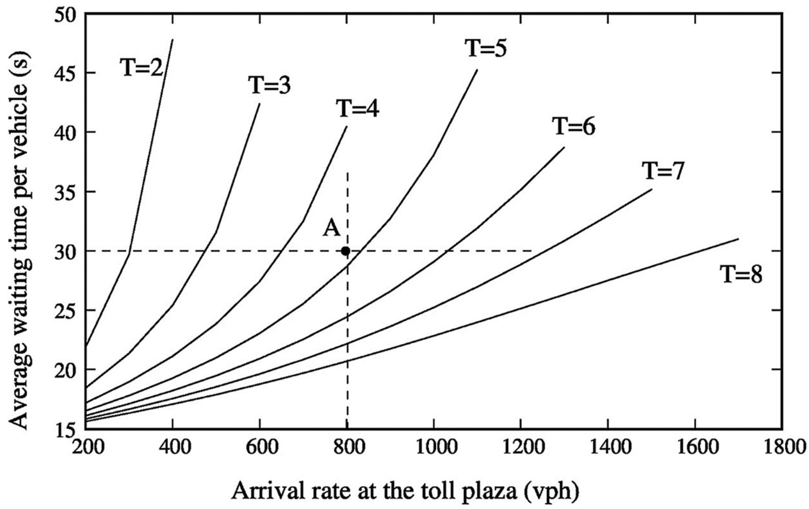

Figure 2. Chart to determine minimum number of toll-lanes required when using the expected waiting time criterion; $\mu_i = 250$ vph (per toll-lane).

the state probabilities depend on them. As before, $E[W]$ is written as $E[W(\lambda, \mu_i, T)]$. Hence, one has to determine the minimum value of T, say T^*, that satisfies $E[W(\lambda, \mu_i, T)] \leq W$.

Figures 2–6 provide $E[W]$ versus λ plots for different μ_i values. These figures can be used to determine the minimum number of toll-lanes required, T^*, for any given value of W, λ and μ_i. The procedure is as follows: (i) for the given value of μ_i choose the appropriate figure, (ii) draw a line parallel to the abscissa at the value of average waiting time per

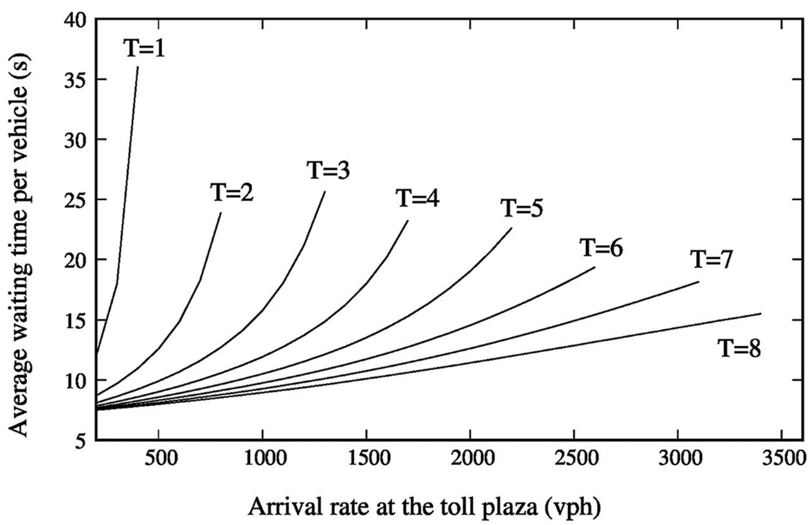

Figure 3. Chart to determine minimum number of toll-lanes required when using the expected waiting time criterion; $\mu_i = 500$ vph (per toll-lane).

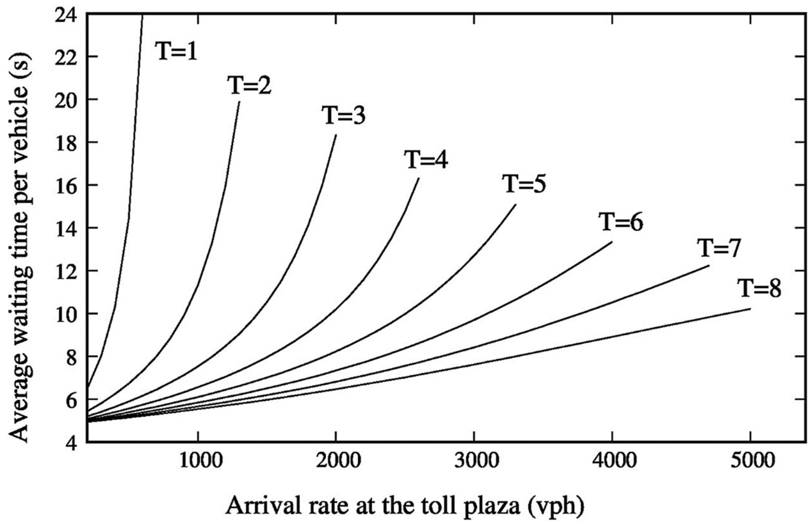

Figure 4. Chart to determine minimum number of toll-lanes required when using the expected waiting time criterion; $\mu_i = 750$ vph (per toll-lane).

vehicle = W, (iii) draw a line parallel to the ordinate axis at the given value of λ; let these two lines intersect at A (iv) locate the plot (of average waiting time per vehicle versus λ) which is immediately to the right of A, and (v) use the T value associated with this plot as T^*. For example, in order to determine the number of toll-lanes for a threshold waiting time of 30 s, arrival rate of 800 vph and service rate of 250 vph (per toll-lane), the process of determining A is shown in Figure 2. Hence, for this case, from Figure 2, T^* is 5.

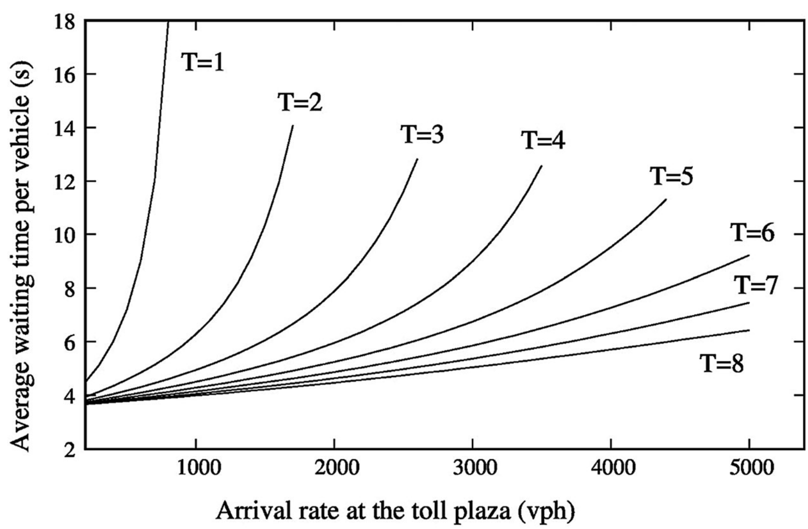

Figure 5. Chart to determine minimum number of toll-lanes required when using the expected waiting time criterion; $\mu_i = 1000$ vph (per toll-lane).

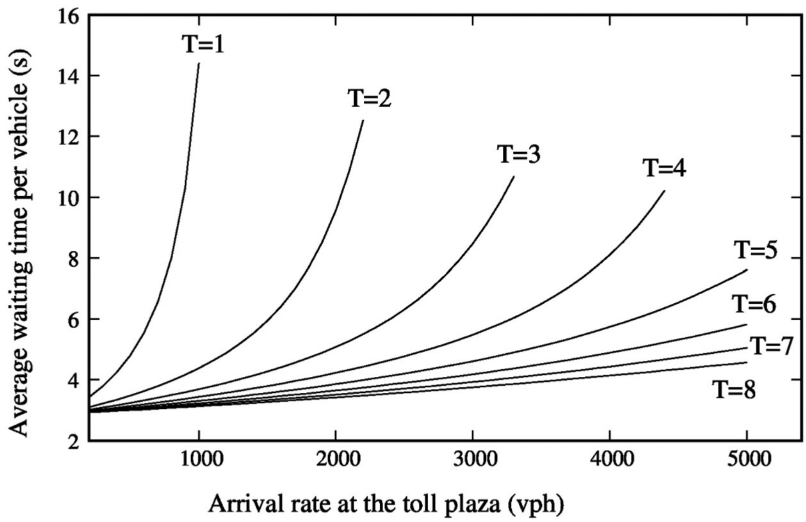

Figure 6. Chart to determine minimum number of toll-lanes required when using the expected waiting time criterion; $\mu_i = 1250$ vph (per toll-lane).

4.2.3. Determination of the length of upstream queueing area using the maximum queue length criterion

The process of determining $\Omega(\lambda, \mu_i, T, Q)$ was explained (see Equation (10)) in the previous section. In that section, given values of λ, μ_i, Q, and threshold probability α, the minimum value of T that satisfied $\Omega(\lambda, \mu_i, T, Q) \leq \alpha$ was determined as T^*. Another way of using the information on $\Omega(\lambda, \mu_i, T, Q)$ is to determine the minimum value of Q, say Q^* that satisfies $\Omega(\lambda, \mu_i, T, Q) \leq \alpha$ for given values of λ, μ_i, T and α. The practical significance of knowing Q^* arises when the number of toll-lanes are fixed or cannot be

Table 7. Minimum length of the toll-lanes when using the maximum queue length criterion; $\mu_i = 250$ vph (per toll-lane).

	λ (vph)				
T	250	500	750	1000	1250
3	3	5	–	–	–
4	2	4	6	–	–
5	2	3	5	7	–
6	2	3	4	6	7
7	2	3	4	5	6
8	2	3	3	4	5

Table 8. Minimum length of the toll-lanes when using the maximum queue length criterion; $\mu_i = 500$ vph (per toll-lane).

	λ (vph)						
T	250	500	750	1000	1500	2000	2500
3	2	3	4	5	–	–	–
4	2	2	3	4	6	–	–
5	1	2	3	3	5	7	–
6	1	2	3	3	4	6	7
7	1	2	2	3	4	5	6
8	1	2	2	3	3	4	5

Table 9. Minimum length of the toll-lanes when using the maximum queue length criterion; μ_i = 750 vph (per toll-lane).

	λ (vph)						
T	500	750	1000	1500	2000	2500	3500
3	2	3	3	5	11	–	–
4	2	2	3	4	5	8	–
5	2	2	3	3	4	5	–
6	2	2	2	3	4	5	7
7	2	2	2	3	3	4	5
8	1	2	2	3	3	4	5

Table 10. Minimum length of the toll-lanes when using the maximum queue length criterion; μ_i = 1000 vph (per toll-lane).

	λ (vph)						
T	500	750	1000	1500	2500	3500	4500
3	2	2	3	4	9	–	–
4	2	2	2	3	5	10	–
5	1	2	2	3	4	6	9
6	1	2	2	3	4	5	6
7	1	2	2	2	3	4	5
8	1	2	2	2	3	4	5

increased beyond a certain value due to limited space and/or other resource constraints. In such situations, or even otherwise, one may have to determine the minimum length of the upstream queueing area (see Figure 1) so that chance of stopped vehicles extending (or overflowing) into the travel lanes is below a certain threshold. The value of Q^* provides the required minimum length when the threshold overflow probability is α.

Values of Q^* for different values of λ, μ_i and T and $\alpha = 0.05$ are provided in Tables 7–11. For example, for a service rate of 750 vph (per toll-lane), arrival rate of 2000 vph and a T value of 5, the required minimum length of the toll-lanes, Q^*, can be read from Table 9 as 4 vehicle-lengths. In order to convert Q^* in vehicle-lengths to Q^*_ℓ, in length units, the expression in Equation (12), where p_i and L_i are the proportion of vehicle type i in the stream and the length of vehicle type i, respectively, can be used. This expression assumes that the fraction of different types of vehicles in any queue is equal to their corresponding share in the traffic stream of the expressway:

$$Q^*_\ell = \left(\sum_{\forall i} p_i L_i \right) Q^*. \tag{12}$$

Table 11. Minimum length of the toll-lanes when using the maximum queue length criterion; μ_i = 1250 vph (per toll-lane).

	λ (vph)						
T	500	750	1000	1500	2500	3500	4500
3	1	2	2	3	5	–	–
4	1	2	2	3	4	6	11
5	1	2	2	2	3	5	6
6	1	1	2	2	3	4	5
7	1	1	2	2	3	4	4
8	1	1	2	2	3	3	4

5. Conclusions

Traffic flow at a toll plaza needs to be studied in terms of how queues develop and dissipate. However, there are no realistic queueing theory-based models of a toll plaza. Attempts have been made in the past; but these attempts either make incorrect assumptions of the queue behaviour (like using multiple server queueing models, such as M/M/S, that tacitly assume a single queue) or are restricted to analysis of toll plazas with only two toll-lanes. The reason for this could possibly be that an analytical model of the queueing observed at a toll plaza, and referred to here as a CMQ2S, is difficult to solve.

The queueing model of a toll plaza proposed here (CMQ2S) incorporates the fact that there are multiple parallel queues at a toll plaza and that arrival to a queue (toll-lane) is dependent on the queue lengths of all the queues. The paper has implemented a PSA-based solution strategy to solve the governing equations of a CMQ2S. The solution provided probability distributions for the queues at the toll-lanes of a toll plaza. The results obtained in this paper were validated using analytical solutions of some special cases of the CMQ2S and simulation results from various scenarios using an application developed in VISSIM (a commercially available micro-simulation model). These validation studies showed that the probability distributions obtained using the PSA-based solution methodology of CMQ2S are reliable.

The paper also showed how the distributions obtained here can be used to design a more efficient toll plaza. For example, the probabilities can be used to gain an insight into how changes in service rate affects the waiting time or the queue lengths. Such an understanding can help decide the level of service that is required at a particular toll plaza. The distributions can also be used to decide how many toll-booths are required at a plaza in order to provide a certain level of service. Decisions on the size of the upstream queueing area, another critical design parameter of a toll plaza, can also be aided by the analysis given here. One can, for instance, determine how long should such queueing areas be so that queues developing at the toll plaza do not overflow into the approach road. In summary, the proposed analysis provides the transportation engineer with a tool that can be used to obtain and evaluate various alternative designs for a toll plaza.

Although the paper presents a rare attempt to rigorously analyse the queueing at a multiple (more than two) toll-lane toll plaza, there are two areas that require improvements. First, the computer implementation of the solution method needs to be made more efficient so that toll plazas with more than eight toll-lanes can also be analysed with reasonable levels of accuracy and computation resources. Second, results with more than one type of toll-lane need to be obtained; although the CMQ2S model proposed here can handle more than one type of toll-lane, numerical results could not be obtained due to lack of multinomial choice models for driver behaviour at toll plazas with more than one type of toll-lane.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Al-Deek, H. M., A. A. Mohamed, and E. A. Radwan. 2000. "New Model for the Evaluation of Traffic Operations at Electronic Toll Collection Plazas." *Transportation Research Record* 1710: 1–10.

Ben-Akiva, M., and S. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge: MIT Press.

Blanc, J. P. C. 1987. "On a Numerical Method for Calculating State Probabilities for Queueing Systems with More than One Waiting Line." *Journal of Computational and Applied Mathematics* 20: 119–125.

Blanc, J. P. C. 1992. "The Power-series Algorithm Applied to the Shortest-Queue Model." *Operations Research* 40 (1): 157–167.

Blanc, J. P. C. 1993. "Performance Analysis and Optimization with Power-Series Algorithm." *Lecture Notes in Computer Science* 729: 53–80.

Conolly, B. W. 1984. "The Autostrada Queueing Problem." *Journal of Applied Probability* 21: 394–403.

Correa, E., C. Metzner, and N. Nino. 2004. "TollSim: Simulation and Evaluation of Toll Stations." *International Transactions in Operational Research* 11 (2): 121–138.

Dubedi, A., P. Chakroborty, D. Kundu, and K. H. Reddy. 2012. "Modelling Automobile Driver's Toll-lane Choice Behaviour at a Toll Plaza." *Journal of Transportation Engineering* 138 (11): 1–8.

Edie, L. C. 1954. "Traffic Delays at Toll Booths." *Journal of the Operations Research Society of America* 2 (2): 107–138.

Gulewicz, V., and J. Danko. 1995. "Simulation-based Approach to Evaluating Optimal Lane-staffing Requirements for Toll Plazas." *Transportation Research Record* 1484: 33–39.

Haight, F. A. 1958. "Two Queues in Parallel." *Biometrika* 45: 401–410.

Hooghiemstra, G., M. Keane, and S. Van De Ree. 1988. "Power Series for Stationary Distributions of Coupled Processor Models." *SIAM Journal of Applied Mathematics* 48 (5): 1159–1166.

Kim, S. 2009. "The Toll Plaza Optimization Problem: Design, Operations and Strategies." *Transportation Research Part E* 45: 125–137.

Mudigonda, S., B. Bartin, and K. Ozbay. 2009. "Microscopic Modeling of Lane Selection and Lane Changing at Toll Plazas." Proceedings of the 88th Annual Meeting of the Transportation Research Board, January, Washington, DC.

Redding, R. T., and A. J. Junga. 1992. "TPASS – Dynamic, Discrete-event Simulation and Animation of a Toll-plaza." Proceedings of the 1992 Winter Simulation Conference, J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, eds., ACM New York, NY, 1292–1295.

Russo, C., R. Harb, and E. Radwan. 2010. "Calibration and Verification of SHAKER, a Deterministic Toll Plaza Simulation Model." *Journal of Transportation Engineering* 136 (2): 85–92.

Schwartz, B. L. 1972. "Queueing Models with Lane Selection: A New Class of Problems." *Operations Research* 22 (2): 331–339.

Taylor, H. M., and S. Karlin. 1984. *An Introduction to Stochastic Modeling*. Orlando, FL: Academic Press.

Thurstone, L. L. 1927. "A Law of Comparative Judgement." *Psychological Review* 34 (4): 273–286.