Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 195 (2009) 127–138
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Stochastics and Statistics

Modeling traffic flow interrupted by incidents

M. Baykal-Gürsoy *, W. Xiao, K. Ozbay

Industrial and Systems Engineering Department, Rutgers University, 96 Frelinghuysen Road, Piscataway, NJ 08854-8018, United States

Received 11 December 2006; accepted 17 January 2008
Available online 26 January 2008

Abstract

A steady-state M/M/c queueing system under batch service interruptions is introduced to model the traffic flow on a roadway link subject to incidents. When a traffic incident happens, either all lanes or part of a lane is closed to the traffic. As such, we model these interruptions either as complete service disruptions where none of the servers work or partial failures where servers work at a reduced service rate. We analyze this system in steady-state and present a scheme to obtain the stationary number of vehicles on a link. For those links with large c values, the closed-form solution of M/M/∞ queues under batch service interruptions can be used as an approximation. We present simulation results that show the validity of the queueing models in the computation of average travel times.
© 2008 Elsevier B.V. All rights reserved.

Keywords: Applied probability; Queueing; Markov modulated

1. Introduction

Increased traffic flow on roadways results in congestion. Congestion leads to delays, decreasing flow rate, higher fuel consumption and thus has negative environmental effects. The cost of total delay in rural and urban areas is estimated by the USDOT to be around $1 trillion per year [27]. Researchers from widely varying disciplines have been paying more and more attention to modeling the vehicular travel in order to improve the efficiency of the current highway systems. The arrival process in roadway traffic is modeled as singly arriving Poisson process [11,45], and as platoons to represent the behavior of the vehicles moving between traffic signals [1,9,13,22]. Daganzo [9] presented a cell transmission model, representing the traffic on a highway with a single entrance and exit, which can be used to predict the evolution of traffic over time and space. Cheah and Smith [8] explored the generality and usefulness of state-dependent M/G/c/c queueing models for modeling pedestrian traffic flows. Heidemann [16] used M/M/1 and M/G/1 queues to model the uninterrupted traffic flow. In order to account for congestion, Jain and Smith [18] used M/G/c/c state-dependent queueing models for modeling and analyzing vehicular traffic flow on a roadway segment which can accommodate a finite number of vehicles. Each vehicle-space corresponds to a server, thus, the maximum number of vehicles that can be accommodated on the link provides the number of servers, c, in the queueing model. Although there are several different types of vehicles utilizing the roadway, in [18] they are all assumed to be identical and considered as a passenger car equivalent (see e.g. [2,46]). Here, the service time gives the total travel time on the link. In this model, service rate (similarly the vehicular traveling speed) is assumed to be a decreasing function of number of vehicles on the link to represent the congestion caused by traffic volume in practice (details of the model can be found in [18]). Heidemann [17] studied transient behavior of M/M/1 queue to analyze the non-stationary traffic flow. Vandaele et al. [47] used M/M/1, M/G/1 and GI/G/1 queues with or without state-dependent rates to model traffic flow. Note that in single-server queueing models each link is considered as a point queue (or vertical queue, see [10]). While in the multi-server case, a link is separated into cells, contrary to the cell transmission model, there is no

* Corresponding author. Tel./fax: +1 732 445 5465.
E-mail address: gursoy@rci.rutgers.edu (M. Baykal-Gürsoy).

0377-2217/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2008.01.024

interdependence between the service times. Van Woensel and Vandaele [48], and Van Woensel et al. [49] validate the use of queueing models empirically and via simulation, respectively. They conclude that M/G/1 queueing models are best to describe the normal traffic flow on a highway, while state-dependent GI/G/m queues were more realistic for the congested traffic. All these models ignore the impact of incidents on the traffic flow.

However, the recurrent congestion generated by excess demand is only part of the problem. Congestion is also caused by irregular occurrences, such as traffic accidents, vehicle disablements, and spilled loads and hazardous materials. An incident is defined here as any occurrence that affects capacity of the roadway [43]. Well over half of nonrecurring traffic delay in urban areas and almost 100% in rural areas are attributed to incidents [27]. The likelihood of secondary incidents increases with the amount of time it takes to clear the initial incident. USDOT estimates that the crashes that result from other incidents make up 14–18% of all crashes [27]. Continuous monitoring of the impact of incidents, and effective incident management can decrease secondary crashes, improve roadway safety and decrease traffic delays.

It is widely accepted that the negative impact of incidents can be significantly reduced by the proper use of incident management procedures. Incident management is the combination of policies and strategies that provide services to reduce the overall incident clearance duration, including incident detection [23,40]. A recent paper by Sheu [41] presents a vehicular-platoon control methodology for automated highway systems in response to lane-blocking incidents.

To improve the efficiency of incident management, mathematical programming methods have been used [34]. Zografos et al. [52] used a districting model to obtain optimal locations of emergency response units to minimize the average incident response time. Daskin [12] constructed a mixed integer programming (MIP) model for simultaneously determining the location, dispatching rule and routing of incident response units. Pal and Sinha [35] also used an MIP model to determine optimal locations for response units that minimize annual cost. Sherali and Subramanian [39] consider the problem of optimal dispatching of response units when future opportunity costs are also taken into account. However, all these models have failed to consider the impact of incidents on the traffic pattern. For instance, the travel time of each link is generally assumed to be constant even during an incident. On the other hand, traffic simulation software packages could be used to analyze the impact of incidents. Many microscopic traffic simulation software packages are applicable to this purpose, for example, Paramics™, INTEGRATION™ and AIMSUN™. Simulation models are also developed to evaluate the performance of various incident management strategies [24,32,36]. Unfortunately, traffic simulation is very time-consuming, due to the need for many replications to reduce the variance and obtain reliable results. The other shortcomings of this approach are:

1. Lack of generality, flexibility and accuracy.
2. May need expensive software, trained personnel and expensive maintenance plans.

Therefore, it is computationally expensive to integrate the traffic simulation software into an incident management evaluation application. There may also be difficulties in integrating the traffic simulation software seamlessly into the whole evaluation system.

In this paper, we analyze the vehicular traffic flow interrupted by incidents using queueing models. Consider vehicles arriving as a Poisson process on a roadway link as shown in Fig. 1, which is subject to traffic incidents. The space occupied by an individual vehicle on the road segment can be considered as one "server", which starts service as soon as a vehicle joins the link and carries the "service" (the act of traveling) until the end of the link is reached [18]. In [18], the number of servers, c, is given as the multiplication of jam density (veh/mi-lane), length of the road link (miles), and number of lanes. We assume service times are exponential and incidents occur randomly with exponential interarrival times. During an incident, the traffic deteriorates such that both the number of working servers and the service rate of all servers decrease. In this state, no new interruptions can arrive. Thus, we are not modeling secondary accidents and instead we focus on primary accidents. As soon as the incident is reported, the incident management system sends a traffic restoration unit to clear the site. The number of working servers and service rates of all servers are restored to their normal level once the incident is cleared that is assumed to take exponential amount of time. Hence, the incident process is modeled as a two-state Markov process representing the normal and incident conditions of the roadway. It is worth to note that the negative impact of incidents involves the reduction of both speed and road capacity. In this study, a lower service rate, $\mu' \geqslant 0$, affecting every

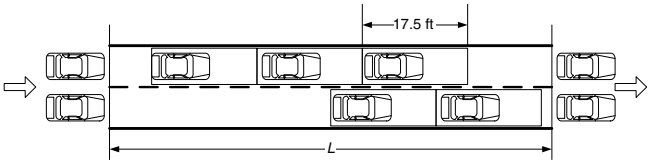

Fig. 1. A two-lane roadway link.

server will be used to represent the impact of congestion caused by incidents. The concepts in this paper also cover, for $c = 1$, the M/M/1 queueing model considered in [16,17,47], when incidents happen disrupting the traffic flow.

We would like to emphasize that the Poisson assumption for vehicle arrivals [48,49], and exponential interarrival times for the incidents [42], are shown to be reasonable. Although the exponential service times may seem unrealistic, we will see later that in our setting, the total time to traverse a link is not going to be exponential. Thus, our model may be considered as having a generally distributed service time.

In the next section, we present the M/M/c queueing model under system level service interruptions to describe the traffic flow on a roadway link that is subject to incidents. Such queues are called M/MSP/c queue, to represent the Markovian behavior of the Service Process (MSP). Two cases of service interruptions will be considered. One is the complete road closure case, i.e., $\mu' = 0$ and the other is the partial failure case, i.e., $\mu' > 0$. Randomly occurring system breakdowns where $\mu' = 0$, have been considered by several researchers. White and Christe [50] studied a single-server queue with preemptive resume discipline, and related such queues to queues with random server breakdowns. Gaver [15] and Keilson [20] also studied a single-server queue with random breakdowns. Gaver [15] obtained the generating functions for the stationary waiting time and the number in the system in an M/G/1 queue. Avi-Itzhak and Naor [3] derived the expected queue length for M/G/1 queue with server breakdowns. Mitrany and Avi-Itzhak [26] analyzed M/M/c queue where each server may be down independently of the others for an exponential amount of time. They obtained an explicit form of the moment generating function of the queue size for one-server and two-server systems, and gave a computational procedure for cases with more than two-servers. Jayawardene and Kella [19] studied an M/G/∞ queue with alternating renewal breakdowns, and they show that the decomposition property holds: the stationary number of customers in the system can be interpreted as the sum of the state of the corresponding system with no interruptions and another nonnegative discrete random variable.

Considering also the partial failure case, Eisen and Tainier [14], Yechiali and Naor [51], and Purdue [37] analyzed the M/M/1 system with two-state Markov modulated service and arrival processes via generating functions. Neuts [29] studied such systems in the context of queues in random environment using matrix–geometric computational methods. Neuts [29,30] also introduced briefly the M/M/c queue in random environment. O'Cinneide and Purdue [31] considered the M/MSP/∞ queue via matrix–algebraic methods and demonstrated with examples the impossibility of "matrix–Poisson" stationary distribution. Keilson and Servi [21] studied the same system and obtained the generating function of the stationary number of customers in the system in terms of Kummer functions [44]. Recently, Baykal-Gürsoy and Xiao [5] considered the M/MSP/∞ system with two-state Markov modulated service process. They showed that this model exhibits the stochastic decomposition property, and gave the explicit form of the stationary distribution.

In fact, there has been a recent interest in these types of queues where the service rate changes randomly (see [6,7,25]). These papers only consider the single-server queue where the asymptotic analysis is performed. In [6,7], the failure times are assumed to be generally distributed. The motivation for these queues can be found in the integrated services communication networks; when web servers serve multi-class customers the secondary customers' service time goes through partial service interruptions since when a high priority customer arrives it receives part of the bandwidth, thus reducing the service speed.

In Section 3, solutions to special cases are presented. In Section 4, we discuss how to approximate the M/MSP/c model when c is large and other computational issues. This paper concludes with a summary of results and recommendations for future research.

2. Notation and mathematical model

Consider a road link as shown in Fig. 1 with c servers that are subject to random interruptions of exponentially distributed durations. Note that in the literature the incident duration is assumed to be normally or log-normally distributed. Here, we assume it is exponentially distributed to provide us with an analytically tractable yet suitable approximation. Later on other models with more general incident durations can be built upon this model. We assume that there is an infinite-capacity buffer space available at the end of the link so that the vehicles that cannot get a server can wait for service. Service times are assumed to be independent and identically distributed exponentials with rate μ. During interruptions, the number of functioning servers decreases from c to c' and the service rates of all servers drop from μ to $\mu' \geqslant 0$. As soon as the interruption is cleared, the number of working servers and their service rates are restored to c and μ, respectively. We assume that interruptions arrive according to a Poisson process with rate f, and the clearance times are i.i.d. exponentials with rate r. The vehicle arrivals are in accordance with a homogeneous Poisson process with intensity λ. Note that we are not considering the traffic during peak hours, where the arrival of the vehicles is uniformly distributed. The interruption and vehicle arrival processes, and the service and clearance times are all assumed to be mutually independent.

The stochastic process $\{X(t), Y(t)\}$ describes the state of the link at time t, where $X(t)$ is the number of vehicles on the link at t, and $Y(t)$ is the status of the link. If at time t, the link is experiencing an interruption, then $Y(t)$ is equal to F (*failure*); otherwise, $Y(t)$ is N (*normal*). The link is said to be in state (i, F), if there are i vehicles on the link which is experiencing

an interruption, while the link is said to be in state (i,N), if there are i vehicles on the link which is functioning normally. Accordingly, we denote the steady-state probability of the link being in state (i,F) by $P_{i,F}$, and the steady-state probability of the link being in state (i,N) by $P_{i,N}$.

The steady-state balance equations are given in (1)–(4).

$0 < i < c'$:

$$\begin{aligned}(\lambda + i\mu' + r)P_{i,F} &= (i+1)\mu' P_{i+1,F} + \lambda P_{i-1,F} + fP_{i,N},\\ (\lambda + i\mu + f)P_{i,N} &= (i+1)\mu P_{i+1,N} + \lambda P_{i-1,N} + rP_{i,F},\end{aligned} \tag{1}$$

$c' \leqslant i < c$:

$$\begin{aligned}(\lambda + c'\mu' + r)P_{i,F} &= c'\mu' P_{i+1,F} + \lambda P_{i-1,F} + fP_{i,N},\\ (\lambda + i\mu + f)P_{i,N} &= (i+1)\mu P_{i+1,N} + \lambda P_{i-1,N} + rP_{i,F},\end{aligned} \tag{2}$$

$c \leqslant i$:

$$\begin{aligned}(\lambda + c'\mu' + r)P_{i,F} &= c'\mu' P_{i+1,F} + \lambda P_{i-1,F} + fP_{i,N},\\ (\lambda + c\mu + f)P_{i,N} &= c\mu P_{i+1,N} + \lambda P_{i-1,N} + rP_{i,F},\end{aligned} \tag{3}$$

$i = 0$:

$$\begin{aligned}(\lambda + r)P_{0,F} &= \mu' P_{1,F} + fP_{0,N},\\ (\lambda + f)P_{0,N} &= \mu P_{1,N} + rP_{0,F}.\end{aligned} \tag{4}$$

Let $G_N(z) = \sum_{i=0}^{\infty} z^i P_{i,N}$ and $G_F(z) = \sum_{i=0}^{\infty} z^i P_{i,F}$, for $|z| \leqslant 1$. Then the generating function of the steady-state number of vehicles on the link is given by

$$G(z) = G_F(z) + G_N(z). \tag{5}$$

By definition, $G(1) = \sum_{i=0}^{\infty} P_{i,N} + \sum_{i=0}^{\infty} P_{i,F} = 1$. Multiplying both sides of (1)–(4) by z^i and summing over all i yield

$$\left[\lambda(1-z) + r + c'\mu'\left(1 - \frac{1}{z}\right)\right] G_F(z) - fG_N(z) = \sum_{i=0}^{c'-1}\left[\left(1 - \frac{1}{z}\right)(c'-i)\mu' z^i P_{i,F}\right] \tag{6}$$

and

$$\left[\lambda(1-z) + f + c\mu\left(1 - \frac{1}{z}\right)\right] G_N(z) - rG_F(z) = \sum_{i=0}^{c-1}\left[\left(1 - \frac{1}{z}\right)(c-i)\mu z^i P_{i,N}\right]. \tag{7}$$

There are $(c + c')$ unknown probabilities in (6) and (7), which could be reduced to two unknowns by using the relations expressed in (1), (2) and (4). Without loss of generality, we assume the remaining unknown probabilities are $P_{0,N}$ and $P_{0,F}$.

Let

$$\underline{g}(z) = \begin{bmatrix} G_N(z) \\ G_F(z) \end{bmatrix},$$

$$A(z) = \begin{bmatrix} \lambda(1-z) + f + c\mu\left(1 - \frac{1}{z}\right) & -r \\ -f & \lambda(1-z) + r + c'\mu'\left(1 - \frac{1}{z}\right) \end{bmatrix},$$

$$b_N(z) = \sum_{i=0}^{c-1}\left[\left(1 - \frac{1}{z}\right)(c-i)\mu z^i P_{i,N}\right],$$

$$b_F(z) = \sum_{i=0}^{c'-1}\left[\left(1 - \frac{1}{z}\right)(c'-i)\mu' z^i P_{i,F}\right]$$

and

$$\underline{b}(z) = \begin{bmatrix} b_N(z) \\ b_F(z) \end{bmatrix},$$

then, (6) and (7) can be rewritten into matrix form, as

$$A(z)\underline{g}(z) = \underline{b}(z). \tag{8}$$

So that $g(z)$ can be obtained by inverting the $A(z)$ matrix. Further manipulations give $G(z)$ as

$$G(z)=\frac{[\lambda z(1-z)+c'\mu'(z-1)+(r+f)z]\sum_{i=0}^{c-1}(c-i)\mu z^i P_{i,N}+[\lambda z(1-z)+c\mu(z-1)+(r+f)z]\sum_{i=0}^{c'-1}(c'-i)\mu' z^i P_{i,F}}{\lambda^2 z^3-(\lambda^2+c\lambda\mu+\lambda f+c'\lambda\mu'+\lambda r)z^2+(c\lambda\mu+c'\lambda\mu'+cc'\mu\mu'+c'f\mu'+c\mu r)z-cc'\mu\mu'} \tag{9}$$

Using the fact $G(1)=1$, and Eqs. (6) and (7), we have

$$\sum_{i=0}^{c'-1}(c'-i)\mu' P_{i,F}+\sum_{i=0}^{c-1}(c-i)\mu P_{i,N}=\frac{f(c'\mu'-\lambda)+r(c\mu-\lambda)}{r+f}. \tag{10}$$

We can immediately see the following stability condition for general multi-server queues.

Stability conditions: (a) The general multi-server queue with exponential service times and batch partial failures is stable if

$$\lambda<\frac{r}{r+f}c\mu+\frac{f}{r+f}c'\mu'.$$

(b) An M/M/c queue with batch system breakdowns is stable if $\frac{\lambda}{c\mu}<\frac{r}{r+f}$.

Besides Eq. (10), we need another equation to solve Eq. (9), which can be deduced from the properties of analytic functions [38].

Let $d(z)$ denote the denominator of $G(z)$.

Remark 1. $d(z)$ has only one root inside the unit circle. In fact, similar to [3] one can show that $d(z)$ has one real root in $(0,1)$, and the other two roots are larger than 1 and also are real. Notice that for any real number z that satisfies $|z|\leqslant 1$, generating function, $G(z)$ is analytic. Therefore, if z_0 is the root of $d(z)$, which satisfies $0<z_0<1$, z_0 should also be a root of the numerator of $G(z)$

$$G(z_0)=\frac{n(z_0)}{d(z_0)}, \tag{11}$$

where $n(z_0)$ is the numerator of $G(z)$ evaluated at z_0. Thus, we must have

$$n(z_0)=0. \tag{12}$$

Combining Eq. (10) and (12) yields the value of $P_{0,N}$ and $P_{0,F}$. With known $P_{0,N}$ and $P_{0,F}$, generating function can be finally obtained by solving (8). Subsequently, the expected number of vehicles on the link is given by evaluating $G'(z)|_{z=1}$.

As an illustrative example, we consider an M/M/5 queue subject to interruptions that reduce the service rate to a third of its normal value. The expected number of vehicles on the link versus the service rate μ is plotted in Fig. 2. In this figure, $\lambda=0.6$, $\mu=3\mu'$, and f and r take some particular values. It can be seen from Fig. 2 that the number of vehicles on the link decreases as service rate increases. If service rate does not change, higher incident frequency or slower clearance rate would lead to more vehicles on the link. Clearly, the stationary number of vehicles on the link when no incident occurs, will constitute the lower bound.

In the following sections we consider:

1. Special cases: $c=1$, single-server system and $c=2$, two-server system.
2. $c\to\infty$. For c large enough, we can use M/MSP/∞ queue as an approximation.
3. Validation of the model via simulation-based models.

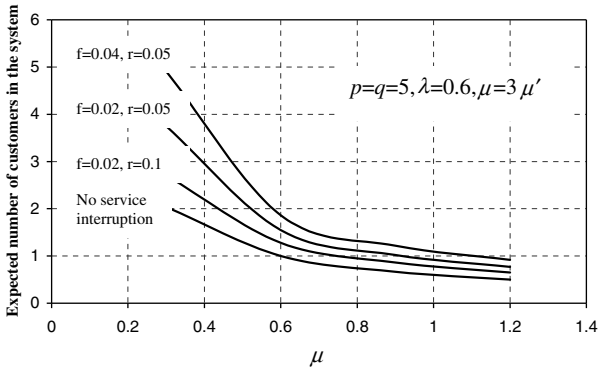

Fig. 2. Expected number of vehicles in an M/M/5 queue subject to service interruptions.

3. Special cases

Since M/M/1 queue in random environment had been studied extensively, we will only briefly discuss these queues with complete breakdown, i.e., $\mu' = 0$, before we consider the M/MSP/2 queue. The generating function is obtained from Eq. (9) as

$$G(z) = \frac{(r\mu - (r+f)\lambda)(-\lambda z + \lambda + r + f)}{(r+f)(\lambda^2 z^2 - \lambda(\lambda + r + f + \mu)z + \mu(\lambda + r))} = \frac{\frac{r}{r+f}(1 - \rho\frac{r+f}{r})(1 - \lambda z/\delta)}{\left[(1-\rho z)(1-\lambda/\delta^z) - \frac{f}{\delta}\right]}, \tag{13}$$

where $\rho = \frac{\lambda}{c\mu}$, with $c = 1$ and $\delta = \lambda + r + f$.

This result matches with the result of Mitrany and Avi-Ithzak [26], who considered an M/M/c queue where each server can breakdown independently of the others. When $c = 1$, the proposed system becomes equivalent to their model. The result also agrees with the general formula derived for the preemptive resume M/G/1 queue [15]. We see that this system does not have the stochastic decomposition property.

The expected number of vehicles on the link can be obtained immediately as (cf. [26]):

$$E[X] = \frac{\lambda[(r+f)^2 + \mu f]}{(r+f)(r(\mu - \lambda) - \lambda f)}. \tag{14}$$

Using Little's formula we obtain the average travel time on the link as

$$W = \frac{[(r+f)^2 + \mu f]}{(r+f)(r(\mu - \lambda) - \lambda f)}. \tag{15}$$

3.1. M/M/2 system subject to interruptions $\mu' > 0$

Consider the two-server system subject to service interruptions. The generating function of this system is given by Eq. (9) as

$$G(z) = \frac{[\lambda z(1-z) - 2\mu'(1-z) + (r+f)z](2\mu P_{0,N} + \mu z P_{1,N}) + [\lambda z(1-z) - 2\mu(1-z) + (r+f)z](2\mu' P_{0,F} + \mu' z P_{1,F})}{\lambda^2 z^3 - (\lambda^2 + 2\lambda\mu + \lambda f + 2\lambda\mu' + \lambda r)z^2 + 2(\lambda\mu + \lambda\mu' + 2\mu\mu' + f\mu' + \mu r)z - 4\mu\mu'}. \tag{16}$$

Using the boundary equation (4) we can rewrite (16) as follows:

$$G(z) = \frac{[\lambda z(1-z) + (r+f)z][(2\mu + \lambda z)P_{0,N} + (2\mu' + \lambda z)P_{0,F}] - 2(1-z)[(\mu'(2\mu + \lambda z) + (\mu' - \mu)fz)P_{0,N} + (\mu(2\mu' + \lambda z) - (\mu' - \mu)rz)P_{0,F}]}{\lambda^2 z^3 - (\lambda^2 + 2\lambda\mu + \lambda f + 2\lambda\mu' + \lambda r)z^2 + 2(\lambda\mu + \lambda\mu' + 2\mu\mu' + f\mu' + \mu r)z - 4\mu\mu'}$$

$G(1) = 1$, or similarly Eq. (10) yields

$$(2\mu + \lambda)P_{0,N} + (2\mu' + \lambda)P_{0,F} = \frac{r(2\mu - \lambda) + f(2\mu' - \lambda)}{r+f}. \tag{17}$$

Moreover, the expected number of vehicles on the link is computed as

$$E(X) = \left[\frac{\mathrm{d}G(z)}{\mathrm{d}z}\right]_{z=1}.$$

In Fig. 3a and b, we plot the expected number of vehicles on the link versus the service rate μ with in the range (1–3) and (0–1), respectively. We let $\mu = 2\mu'$,$\lambda = 1$, and f and r take some particular values. These figures show that the number of vehicles on the link decreases as service rate increases, but the effect is more significant when μ is smaller than 1. Clearly, this is due to the stability condition. Note that, the stability condition requires that $\mu > \frac{(f+r)\lambda}{(f+2r)} = \frac{f+r}{f+2r}$.

Fig. 4 shows the effect of increasing μ' on the number of vehicles on the link, while keeping μ fixed at 2. Similar to Fig. 3, we let $\lambda = 1$, and f and r take some particular values. It can be seen that the expected number of vehicles also decreases as μ' increases.

3.2. M/M/2 system subject to interruptions with $\mu' = 0$

When $\mu' = 0$, we can also obtain the closed-form solution for M/M/2 queues that are subject to interruptions. Letting $\mu' = 0$ in Eqs. (6) and (7) yields:

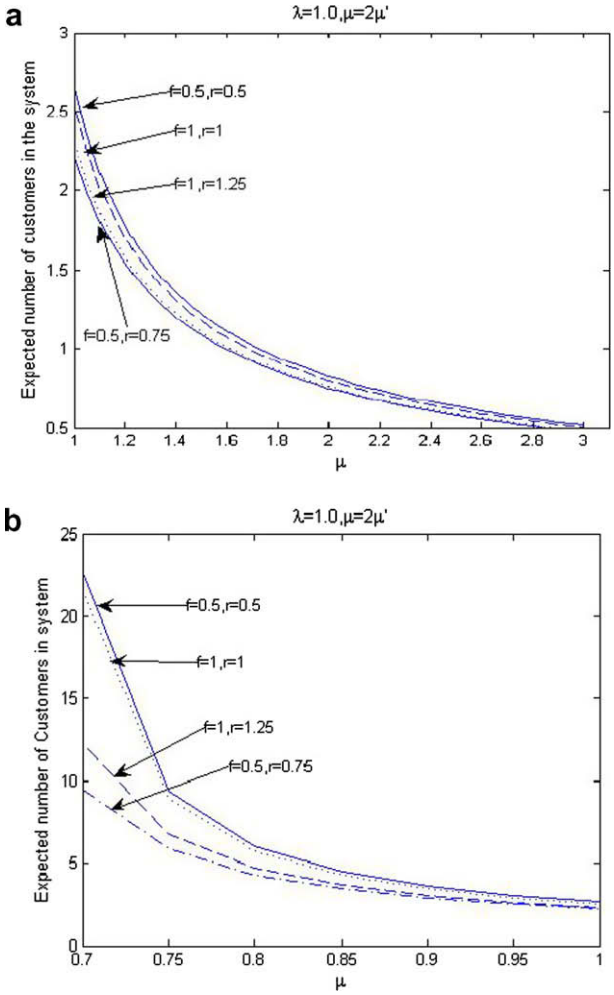

Fig. 3. Expected number of vehicles in an M/MSP/2 queue ($\lambda = 1.0$, $\mu = 2\mu'$).

$$\begin{cases} (\lambda(1-z)+r)G_F(z) - fG_N(z) = 0, \\ (\lambda(1-z)+2\mu(1-1/z)+f)G_N(z) - rG_F(z) = 2\mu(1-1/z)P_{0,N} + \mu z(1-1/z)P_{1,N}. \end{cases}$$

Since $G(z) = G_F(z) + G_N(z)$ and $G(1) = 1$, we have

$$2P_{0,N} + P_{1,N} = \frac{2r}{r+f} - \frac{\lambda}{\mu} = 2\left(\frac{r}{r+f} - \frac{\lambda}{2\mu}\right).$$

With $\mu' = 0$, the boundary condition, $(\lambda + r)P_{0,F} = fP_{0,N}$ and $(\lambda + f)P_{0,N} = \mu P_{1,N} + rP_{0,F}$ imply

$$P_{1,N} = \frac{1}{\mu}\left(\lambda + f - \frac{rf}{\lambda + r}\right)P_{0,N} = \frac{\lambda}{\mu}\left(\frac{\lambda + r + f}{\lambda + r}\right)P_{0,N}.$$

Balance equations together with the previous equality, yields

$$P_{0,N} = \frac{\left(\frac{r}{r+f} - \frac{\lambda}{2\mu}\right)}{1 + \frac{\lambda}{2\mu}\left(\frac{\lambda+r+f}{\lambda+r}\right)}, \tag{18}$$

giving the stability condition, as $\frac{\lambda}{2\mu} < \frac{r}{r+f}$. The generating function is also obtained as

$$G(z) = \frac{\frac{r}{r+f}\left(1 - \rho\frac{r+f}{r}\right)(1 - \lambda z/\delta)}{(1-\rho z)(1-\lambda/\delta^z) - \frac{f}{\delta}} \frac{1+z\eta}{1+\eta}, \tag{19}$$

where $\rho = \frac{\lambda}{2\mu}$, $\eta = \frac{\lambda}{2\mu}\left(\frac{f+\lambda+r}{\lambda+r}\right)$. Clearly, similar to M/MSP/1, the M/MSP/2 system does not exhibit the stochastic decomposition property.

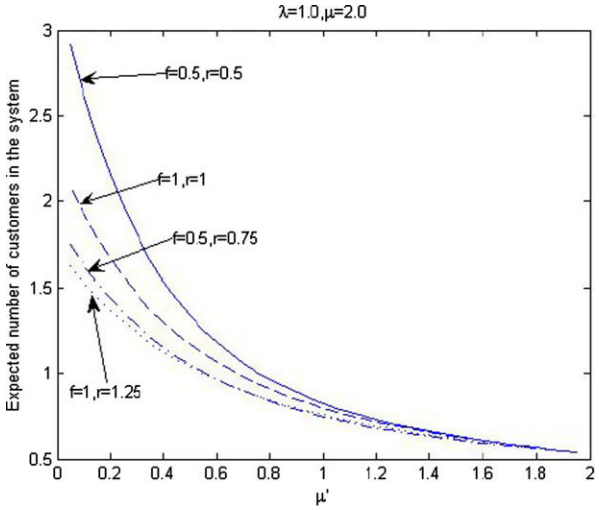

Fig. 4. Expected number of vehicles in an M/MSP/2 queue ($\lambda = 1.0$, $\mu = 2.0$).

The expected number of vehicles on the link is

$$E(X) = \frac{2\lambda\mu[2(\lambda+r)(f^2+r^2+2\mu f)+f(2r+\lambda)^2]}{(f+r)(2r\mu-f\lambda-r\lambda)(f\lambda+(r+\lambda)(\lambda+2\mu))} \tag{20}$$

and using Little's formula we obtain the average travel time on the link as

$$W = \frac{2\mu[2(\lambda+r)(f^2+r^2+2\mu f)+f(2r+\lambda)^2]}{(f+r)(2r\mu-f\lambda-r\lambda)(f\lambda+(r+\lambda)(\lambda+2\mu))}. \tag{21}$$

4. Computational issues

4.1. Approximating M/M/c queue with service interruptions

Generally, a roadway link can accommodate hundreds of vehicles. As we have mentioned in the previous section, when c is large, it is not easy to obtain explicit expressions for the generating function. The computational complexity of M/MSP/c queues for large c values motivates us to seek simpler solutions. A straightforward approach is to use M/MSP/∞ to approximate M/MSP/c if c is large enough.

In our previous work [5], we show that an M/MSP/∞ queueing system exhibits the stochastic decomposition property, namely, the stationary number of vehicles present on the link at a random point in time can be represented as the sum of two independent random variables. One of these is the stationary number of customers present in an ordinary M/M/∞ queue without interruptions. For completeness, we present our result briefly as follows.

If we let $a = \frac{f}{\mu}, b = \left(\frac{f}{\mu}+\frac{r}{\mu'}\right)$, $\rho^* = \frac{1}{2}\left(\frac{\lambda}{\mu}-\frac{\lambda}{\mu'}\right)$ and $p = \frac{r\mu+f\mu'}{r\mu+f\mu}$, then the generating function of the stationary number of customers in an M/MSP/∞ queue can be computed as

$$G(z) = e^{\frac{\lambda}{\mu}(z-1)}(pM(a,b,-2\rho^*(z-1)) + (1-p)M(a+1,b+1,-2\rho^*(z-1))),$$

where $M(a,b,w)$ is the Kummer's function [44]. It can be shown that Kummer functions are the generating functions of Poisson random variables randomized by truncated beta. The expected number of customers in the system is computed, as

$$E(X) = \frac{\lambda}{\mu} + \frac{\lambda f(\mu-\mu')}{\mu^2(r+f)}\left(1+\frac{(f+\mu)(\mu-\mu')}{(r\mu+f\mu'+\mu\mu')}\right). \tag{22}$$

In Fig. 5, we plot the expected number of vehicles on the link versus the service rate μ. We let $\lambda = 6$, $\mu = 10\mu'$, and f and r take some particular values.

To verify this approximation, we compute the expected number of vehicles in an M/MSP/200 queue and compare it with the result of M/MSP/∞ queue. Both queues have the same incident arrival rate, incident clearance rate and vehicle arrival rate, which are 0.002, 0.075 and 6, respectively. In both queues, we assume the service rate during an interruption is 1/10 of the normal service rate. We use balance equations (1)–(4) up to a high enough capacity to obtain the steady-state

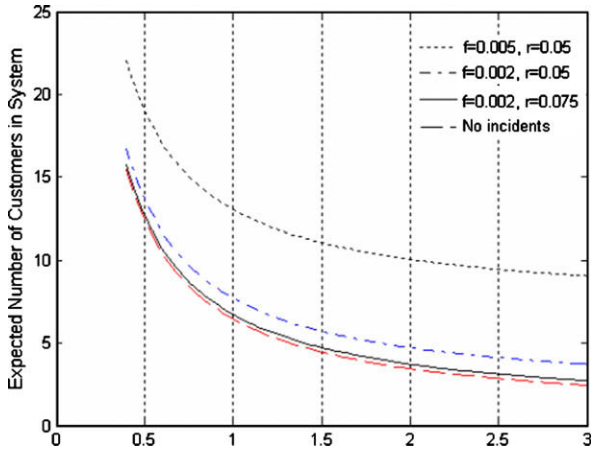

Fig. 5. Expected number of vehicles in an M/MSP/∞ queue with $\lambda = 6$ and $\mu = 10\mu'$.

distribution, so that the blocking probability is negligible (in the order of 10^{-6}). Table 1 gives almost identical results for these queues with service rate varying from 0.3 to 2.7.

To see the effect of loading the system, the arrival rate, λ is increased from 6 to 51. Table 2 summarizes the computational results for M/MSP/200 and M/MSP/∞ queues. Here the blocking probability is negligible at a level less than 0.009. Relative errors defined below are also listed:

$$e_{\text{relative}} = \frac{|E(X)_{\text{M/M}/c} - E(X)_{\text{M/M}/\infty}|}{E(X)_{\text{M/M}/c}} \times 100\%.$$

Note that all the relative errors are less than 9%, which means using M/MSP/∞ is a promising approximation for M/MSP/c with large c value even for a highly loaded system.

4.2. Validation of M/MSP/c and M/MSP/∞ models

Consider a roadway link depicted in Fig. 1. We can use our models to investigate the impact of the incidents on the average travel time. The average travel time can be obtained from the Little's theorem as given as the first equality below. The second equality is only valid for the M/MSP/∞ queue

$$W = E(X)/\lambda = \frac{1}{\mu} + \frac{f(\mu - \mu')}{\mu^2(r+f)}\left(1 + \frac{(f+\mu)(\mu-\mu')}{(r\mu + f\mu' + \mu\mu')}\right). \tag{23}$$

In order to verify our model, we compare our analytical solutions with the results of INTEGRATION™ (version 1.5 × 3), a widely used microscopic traffic simulation software package. To this end, we need to transform the parameters in Eq. (23) into inputs for INTEGRATION™ as follows:

Table 1
The expected number of vehicles in M/MSP/200 and M/MSP/∞

μ	0.3	0.6	0.9	1.5	2.7
M/MSP/200	21.675	11.171	7.588	4.655	2.640
M/MSP/∞	21.675	11.171	7.5884	4.6548	2.6401

$\lambda = 6$ and $\mu = 10\mu'$, $f = 0.002$, $r = 0.075$.

Table 2
The expected number of vehicles in M/MSP/200 and M/MSP/∞

λ	6	12	24	51
M/MSP/200	21.68	43.59	89.18	202.75
M/MSP/∞	21.68	43.35	86.70	184.24
Relative error	0%	0.5%	2.7%	8%

$\mu = 0.3$ and $\mu' = 0.03$, $f = 0.002$, $r = 0.075$.

1. Length of this link, L: According to the ITE (1994), average length of the occupancy of a vehicle is 17.5 feet, thus, the length of a 2-lane link which can accommodate c vehicles is computed as $(c \times 17.5/2) = 8.75c$ feet $= 0.002667c$ km.
2. Traffic demand: Number of vehicles arriving at the link per hour in terms of λ (veh/seconds) $= 3600\lambda$ (veh/hour).
3. Frequency of incidents, f (incident/seconds): We use f as the rate to generate incidents in the simulation.
4. Incident duration: $d = 1/r$ seconds.
5. Travel speed at full capacity: $v = L \times \mu \times 3600$ km/hour.
6. Travel speed during incidents: $v' = L \times \mu' \times 3600$ km/hour.
7. Given the number of lanes on a link, n, the number of lanes blocked by the incident is, $b = n \times (1 - \mu'/\mu)$.

We use INTEGRATION™ to simulate a link with travel speed at full capacity $v = 57.5$ km/hour (65 miles/hour). Note that INTEGRATION™ treats the arrival process as fluid, thus generating λ vehicles per hour deterministically. We consider various arrival rates and link lengths on a two-lane roadway where minor incidents happen. Minor incidents take less than 30 minutes to be cleared [28] with the average of 7 minutes [42]. Skabardonis et al. report 0.5 incidents per hour for a 1 km roadway [42]. The values for f and r are chosen accordingly. For each setting, we run the simulation to obtain average travel times for 100 replications and each replication simulates a 12,000-second period. Under congestion, each replication takes 5 minutes, thus each scenario takes more than 4 hours. This is very time-consuming compared to the analytical model. Tables 3 and 4 summarize the simulation and analytical results. The last column shows the relative errors. We would like to emphasize that in the simulation model as in real life, the service times are also neither independent nor exponential. The incident process is the only random process in the simulations. Still, the restrictive analytical model performs reasonably for obtaining "average" performance measures giving relative errors within 13% range.

Table 3 is used to demonstrate the effect of decreasing c by decreasing L. Note that decreasing the length of the link in simulation leads to the increase in the service rate in the analytical approach, since the full-capacity traveling speed remains constant. The results show that the relative error for both M/MSP/c and M/MSP/∞ are higher for shorter links. If the link is long enough, the relative error can be kept within the range of 2%. This is reasonable enough for most applications.

Table 4 is used to demonstrate the effect of congestion on the approximation. As congestion increases due to increasing arrival rate, relative errors increase. But, clearly, as congestion increases M/MSP/c becomes a better model for the interrupted traffic than the infinite server system. We should note that the traffic intensity changes as the arrival rate increases

Table 3
Comparison of simulation and analytical results (decreasing c)

c (veh)	λ (veh/seconds)	μ (veh/seconds)	$\mu' = \frac{1}{14} \cdot \mu$ (veh/seconds)	f (1/seconds)	r (1/seconds)	Model	Average travel time (seconds)	Relative error (%)
400	0.3	0.015	0.001071	0.0002	0.005	M/M/c	74.5696	0.48
						M/M/∞	74.5696	0.48
						Simulation	74.93	
200	0.3	0.03	0.002143	0.0002	0.005	M/M/c	39.1883	1.88
						M/M/∞	39.1883	1.88
						Simulation	39.94	
100	0.3	0.06	0.004286	0.0002	0.005	M/M/c	20.8397	10.87
						M/M/∞	20.8397	10.87
						Simulation	23.38	
50	0.3	0.12	0.008571	0.0002	0.005	M/M/c	11.0764	12.85
						M/M/∞	11.0761	12.86
						Simulation	12.71	

Table 4
Comparison of simulation and analytical results (increasing λ)

c (veh)	λ (veh/seconds)	μ (veh/seconds)	$\mu' = \frac{1}{14} \cdot \mu$ (veh/seconds)	f (1/seconds)	r (1/seconds)	Model	Average travel time (seconds)	Relative error (%)
200	0.3	0.03	0.002143	0.0002	0.005	M/M/c	39.1883	1.88
						M/M/∞	39.1883	1.88
						Simulation	39.94	
200	0.5	0.03	0.002143	0.0002	0.005	M/M/c	39.2001	10.89
						M/M/∞	39.1883	10.92
						Simulation	43.99	
200	0.7	0.03	0.002143	0.0002	0.005	M/M/c	39.3555	11.89
						M/M/∞	39.1883	12.27
						Simulation	44.67	

from 0.05 to 0.12 during the regular traffic, and from 18.19 to 42.46 during the incidents in Table 4. But, the steady-state traffic intensity is still low, between 0.051 and 0.121.

5. Conclusion and future research

In this paper, we propose queueing models to describe the traffic flow on a road link that is subject to roadway incidents, and we explore their solution schemes. For some special cases, we present closed-form solutions. We also investigate the use of M/MSP/∞ system to approximate an M/MSP/c system when c is large. The comparison of the analytical results with the simulation results via INTEGRATION™, a traffic simulation package, shows that M/MSP/∞ provides a good enough approximation for long links while M/MSP/c might be more appropriate for the congested roadways. This supplies an alternative approach to obtain the average link travel time under the impact of incidents. The amount of time required to run the INTEGRATION™ simulation software increases linearly with the number of replications, while the computation time of our model is fixed. The advantages of the proposed models over the simulation approach are significant when multiple simulation runs are required to reduce the variance in simulation results.

Empirical validation of our model is currently not possible because real-time queuing and delay data due to accidents are not readily available. Since accidents are random events, it is almost impossible to predict their location and time for real-time data collection. Traffic detectors that are, in general, sparsely deployed do not capture the type of delay and queueing information needed to validate our model. However, there are some emerging technologies such as the use of instrumented probe vehicles or cell phones continuously roaming in the network that can be used for our purposes when they become more widely available for real-time data collection. In fact, one of the co-authors of this paper is actively involved in the collection of such real-time data [4,33]. In the next phase of our research, we will try to collect data using one or more of these emerging data collection technologies.

Even though the proposed queueing models are appropriate for obtaining the average performance measures, in the future, we plan to validate these analytical models in terms of their variance characteristics.

Acknowledgements

The authors would like to thank Richard S. Falk for providing a crucial reference. They are grateful to Andrew Ross for noticing an error in the previous version of Table 1, and Zhe Duan for his help in preparing Figs. 3, 4 and Tables 2–4. The authors acknowledge with gratitude the suggestions by anonymous referees that helped to improve the presentation of the paper.

References

[1] A.S. Alfa, M.F. Neuts, Modeling vehicular traffic using the discrete time Markovian arrival process, Transportation Science 29 (2) (1999) 109–117.
[2] A. Al-Kaisy, Y. Jung, H. Rakha, Developing passenger car equivalency factors for heavy vehicles during congestion, Journal of Transportation Engineering 131 (7) (2005) 514–523.
[3] B. Avi-Itzhak, P. Naor, Some queueing problems with the service station subject to breakdown, Operations Research 11 (1963) 303–320.
[4] B. Bartin, K. Ozbay, C. Iyigun, Clustering-based methodology for determining optimal roadway configuration of detectors for travel time estimation, Transportation Research Board: Journal of Transportation Research Record 2000 (2007) 98–105.
[5] M. Baykal-Gürsoy, W. Xiao, Stochastic decomposition in M/M/∞ queues with Markov modulated service rates, Queueing Systems 48 (1–2) (2004) 75–88.
[6] O.J. Boxma, I.A. Kurkova, The M/M/1 queue in a heavy-tailed random environment, Statistica Neerlandica 54 (2) (2000) 221–236.
[7] O.J. Boxma, I.A. Kurkova, The M/G/1 queue with two service speeds, Advances in Applied Probability 33 (2001) 520–540.
[8] J.Y. Cheah, J.M. Smith, Generalized M/G/C/C state dependent queuing models and pedestrian traffic flows, Queueing Systems 15 (1994) 365–385.
[9] C.F. Daganzo, The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory, Transportation Research – Part B 4 (1994) 269–287.
[10] C.F. Daganzo, Fundamentals of Transportation and Traffic Operations, Pergamon–Elsevier, Oxford, UK, 1997.
[11] J.N. Darroch, G.F. Newell, R.W.J. Morris, Queues for vehicle-actuated traffic light, Operations Research 12 (1964) 882–895.
[12] M.S. Daskin, Location, dispatching and routing models for emergency services with stochastic travel times, in: A. Ghosh, G. Rushton (Eds.), Spatial Analysis and Location-Allocation Models, Van Nostrand, 1987, pp. 224–265.
[13] M.C. Dunne, Traffic delays at a signalized intersection with binomial arrivals, Transportation Science 1 (1967) 24–31.
[14] M. Eisen, M. Tainiter, Stochastic variations in queueing processes, Operations Research 11 (6) (1963) 922–927.
[15] D.P. Gaver Jr., A waiting line with interrupted service, including priorities, Journal of the Royal Statistical Society Series B 24 (1962) 73–90.
[16] D. Heidemann, A queueing theory approach to speed–flow–density relationships, in: Proceedings of the 13th International Symposium on Transportation and Traffic Theory, France, July 1996.
[17] D. Heidemann, A queueing theory model of nonstationary traffic flow, Transportation Science 35 (4) (2001) 405–412.
[18] R. Jain, J.M. Smith, Modeling vehicular traffic flow using M/G/C/C state dependent queueing models, Transportation Science 31 (4) (1997) 324–336.
[19] A.K. Jayawardene, O. Kella, M/G/∞ with alternating renewal breakdowns, Queueing Systems 22 (1996) 79–95.
[20] J. Keilson, Queues subject to service interruptions, Annals of Mathematical Statistics 33 (1962) 1314–1322.
[21] J. Keilson, L.D. Servi, The matrix M/M/∞ system: Retrial models and Markov modulated sources, Advances in Applied Probability 25 (1993) 453–471.

[22] P. Lehoczky, Traffic intersection control and zero-switch queues, Journal of Applied Probability 9 (1972) 382–395.
[23] W.-H. Lin, C.F. Daganzo, The simple detection scheme for delay-inducing freeway incidents, Transportation Research – Part A 31 (2) (1997) 141–155.
[24] H. Liu, R.W. Hall, INCISIM: Users Manual. California PATH Research Report, UCB-ITS-PWP-2000-15, 2000.
[25] S.R. Mahabhashyam, N. Gautam, On queues with Markov-modulated service rates, Queueing Systems 51 (1–2) (2005) 89–113.
[26] I.L. Mitrany, B. Avi-Itzhak, A many-server queue with service interruptions, Operations Research 16 (3) (1968) 628–638.
[27] National Conference on Traffic Incident Management: A Road Map to the Future, Proceedings, 2–4 March, 2002.
[28] National Cooperative Highway Research Program, Synthesis 318, Safe and Quick Clearance of Traffic Incidents, TRB, 2003. <http://cms.transportation.org/sites/ssom/docs/NCHRP_syn_318.pdf>.
[29] M.F. Neuts, Matrix–Geometric Solutions in Stochastic Models: An Algorithmic Approach, The John Hopkins University Press, 1981.
[30] M.F. Neuts, Further results on the M/M/1 queue with randomly varying rates, OPSEARCH 15 (4) (1978) 158–168.
[31] C. O'Cinneide, P. Purdue, The M/M/∞ queue in a random environment, Journal of Applied Probability 23 (1986) 175–184.
[32] K. Ozbay, B. Bartin, Evaluation of incident management systems using simulation, Simulation Journal 79 (2) (2003) 69–82.
[33] K. Ozbay, B. Bartin, S. Chien, South Jersey real-time motorist information system: Technology and practice, Journal of Transportation Research Record (1886) (2004) 68–76.
[34] K. Ozbay, P. Kachroo, Incident Management in Intelligent Transportation System, Artech House Books, Boston, 1999.
[35] R. Pal, K. Sinha, A framework for locating highway incident response vehicles in urban areas, INFORMS National Meeting, San Diego, California, 1987.
[36] R. Pal, K. Sinha, Simulation model for evaluating and improving effectiveness of freeway service patrol programs, Journal of Transportation Engineering 128 (4) (2002) 355–365.
[37] P. Purdue, The M/M/1 queue in a Markovian environment, Operations Research 22 (1973) 562–569.
[38] W. Rudin, Real and Complex Analysis, second ed., McGraw-Hill, 1974.
[39] H.D. Sherali, S. Subramanian, Opportunity cost-based models for traffic incident response problem, Journal of Transportation Engineering 125 (3) (1999) 176–185.
[40] J.-B. Sheu, A sequential detection approach to real-time freeway incident detection and characterization, European Journal of Operational Research 157 (2004) 471–485.
[41] J.-B. Sheu, Microscopic modeling and control logic for incident-responsive automatic vehicle movements in single-automated-lane highway systems, European Journal of Operational Research 182 (2007) 640–662.
[42] A. Skabardonis, K.F. Petty, R.L. Bertini, P.P. Varaiya, H. Noeimi, D. Rydzewski, I-880 Field Experiment: Analysis of Incident Data, TRB vol. 1603, Washington, DC, 1997.
[43] A. Skabardonis, K. Petty, P. Varaiya, R. Bertini, Evaluation of the freeway service patrol (FSP) in Los Angeles, UCB-ITS-PRR-98-31, California PATH Research Report, Institute of Transportation Studies, University of California, Berkeley, 1998.
[44] L.J. Slater, Confluent hypergeometric functions, in: M. Abromowitz, I. Stegun (Eds.), Handbook of Mathematical Functions, CUP, Cambridge, UK, 1964, pp. 503–515.
[45] J.C. Tanner, A problem of interface between two queues, Biometrica 40 (1953) 58–69.
[46] Transportation Research Board, 2000. Highway Capacity Manual, fourth ed. National Research Council, Washington, DC.
[47] N. Vandaele, T. Van Woensel, N. Verbruggen, A queueing based traffic flow model, Transportation Research – Part D: Transportation and Environment 5 (2) (2000) 121–135.
[48] T. Van Woensel, N. Vandaele, Empirical validation of a queueing approach to uninterrupted traffic flows, 4OR-A Quarterly Journal of Operations Research 4 (1) (2006) 59–72.
[49] T. Van Woensel, B. Wuyts, N. Vandaele, Validating state-dependent queueing models for uninterrupted traffic flows using simulation, 4OR-A Quarterly Journal of Operations Research 4 (2006) 159–174.
[50] H.C. White, L.S. Christe, Queuing with preemptive priorities or with breakdown, Operations Research 6 (1958) 79–95.
[51] U. Yechiali, P. Naor, Queueing problems with heterogeneous arrivals and service, Operations Research 19 (1971) 722–734.
[52] K.G. Zografos, T. Nathanail, P. Michalopoulos, Analytical framework for minimizing freeway-incident response time, Journal of Transportation Engineering ASCE 119 (1993) 535–549.