M/M/C Queues with Markov Modulated Service Processes

Melike Baykal-Gursoy
Industrial and Systems Eng.
Rutgers, The State Univ. of NJ
96 Frelinghuysen Rd
Piscataway, NJ 08854-8018
gursoy@rci.rutgers.edu

Zhe Duan
Industrial and Systems Eng.
Rutgers, The State Univ. of NJ
96 Frelinghuysen Rd
Piscataway, NJ 08854-8018
zheduan@eden.rutgers.edu

ABSTRACT

Motivated by the need to study traffic flow affected by incidents we consider M/M/C queueing system where servers operate in a Markovian environment. 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 all servers work at some reduced service rate. We analyze the system with multiple failure states in steady state and present a scheme to obtain the stationary number of vehicles on a link. The special case of single breakdown case is further analyzed and performance measures in closed form are obtained.

1. INTRODUCTION

Increased demand for roadway travel on existing roadways results in a rise in congestion. Congestion leads to delays, decreasing flow rate, higher fuel consumption and inevitably pollution, 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 [29]. Researchers from widely varying disciplines have been paying attention to modeling the vehicular travel in order to improve the efficiency of the current highway systems. Classical traffic models are mostly based on the treatment of interacting vehicles, their statistical distribution, or their average velocity and density as a function of time and space. Main modeling approaches can be classified as microscopic (particle-based) (see e.g., Gazis *et al.* [14, 15], May and Keller [27]), mesoscopic (gas-kinetic) (see e.g., Prigogine and Herman [37]), and macroscopic (fluid-dynamic) (see e.g., Lighthill and Whitham [24], and Richards [39]) models (Helbing [19]). An alternative approach is the queueing models that determine the travel times as a function of entering and leaving flows. Initially, queueing analysis has been mainly utilized for the performance evaluation via deterministic models (May and Keller [26], and Newell [34]) for traffic light synchronization (Newell [33]). The stochastic models include M/M/1 and M/G/1 queues considered by Heidemann [17, 18], and Vandaele *et al.*[41], and M/G/C/C state dependent models studied by Cheah and Smith [7],and Jain and Smith [20], where the service rate (the vehicular traveling speed) is assumed to be a decreasing function of the number of the customers in the system to represent the congestion caused by the traffic volume in practice. Although this latter queueing model considers congestion, they all ignore the impact of randomly occuring 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 (Skabardonis *et al.*[40]). Well over half of nonrecurring traffic delay in urban areas and almost 100% in rural areas are attributed to incidents [29]. 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 crashess [29]. Continuous monitoring of the impact of the incident, and effective incident management can decrease secondary crashes, improve roadway safety and decrease traffic delays.

In this paper, we analyze the vehicular traffic flow interrupted by incidents using queueing models. Consider vehicles traveling on a roadway link as shown in Figure 1, which

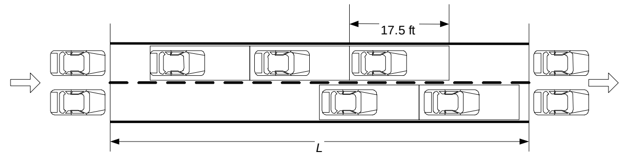

Figure 1: A two-lane roadway link [4]

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" until the end of the link is reached. During an incident, the traffic deteriorates such that the service rate of all servers decrease. Once an incident occurs, the incident management system sends a traffic restoration unit to fix it. The service rates of all servers are

ValueTools'06 , October 14, 2006, Pisa, Italy

restored to their prior level at the clearance of the incident. These system level interruptions and restorations are modeled as a Markovian service process(MSP).

Randomly occurring server breakdowns have been considered for certain queueing models. Researchers studied a single server queue with random server breakdowns (White and Christie [42], Gaver[13], Keilson[22], Avi-Itzhak and Naor [2], Halfin[16], Federgruen and Green [10, 11], and Fischer[12]), M/M/C queue where each server may be down independently of the others for an exponential amount of time (Mitrani and Avi-Itzhak [28]), M/G/∞ queue with alternating renewal breakdowns (Jayawardene and Kella [21]). Jayawardene and Kella [21] show that the decomposition property, a well known property of vacation type queues, holds for such queues: 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, the M/M/1 system in a two-state Markovian environment where the arrival as well as the service process are affected, is analyzed via generating functions first by Eisen and Tainiter [9], then by Yechiali and Naor [44], and Purdue[38]. Such queues, in general, in n-state Markovian environment are said to have Markovian arrival process(MAP)(see, e.g., Neuts [32]) and Markovian service process (MSP), and might be represented in Kendall notation as MAP/MSP/1 [36]. Yechiali[43] considered the general MAP/MSP/1 queue. Neuts [31, 30] studied M/M/1 and briefly M/M/C queues in a random environment using matrix-geometric computational methods. O'Cinneide and Purdue [35], and Keilson and Servi [23] analyzed the n-state MAP/MSP/∞ queue. For all these queueing models no explicit solution was given. For the special case of M/M/∞ queue with two-state Markov modulated arrival process , Keilson and Servi [23] show that the decomposition property holds, and provided the explicit solution.

Recently, Baykal-Gursoy and Xiao [3] considered the M/M/∞ system with the two-state Markov modulated service process, e.g., M/MSP/∞ queue. Using the method introduced in [23], they proved that this model also exhibits a stochastic decomposition property, and gave the explicit form of the stationary distribution. For the infinite server queue with two-state service mechanism, [21] in the complete breakdown case, and [3] also in the partial failure case, are the first papers showing the validity of decomposition property. In fact, there has been a recent interest in the systems where the service rate changes randomly for the single server queue (Adan and Kulkarni [1], Boxma and Kurkova [5, 6], and Mahabhashyam and Gautam [25]), and M/M/C queue with two-state Markov modulated service (Baykal-Gursoy *et al.* [4]). The motivation for such single server queues can be found in the integrated services communication networks (see the references in [5, 6]).

2. MATHEMATICAL MODEL

Consider a road link as shown in Figure 1 with C servers that are subject to random system interruptions of exponentially distributed durations. We assume that there is buffer space available in front of the link so that the vehicles that cannot get a server can wait for service. As the most general case we consider M/M/C queues with n types of server states. The server states are denoted as $S_1, \ldots, S_n$ that have associated service rates $\mu_1, \ldots, \mu_n$ respectively. Service times are assumed to be independent and identically distributed (i.i.d.) exponentials. The vehicle arrivals are in accordance with a homogeneous Poisson process with intensity λ irrespective of the server state. Movements between server states include only the moves to the adjacent states as shown in Figure 2. The state transitions at the boundary states could be presented respectively. This example represents the case where S_1 corresponds to the normal state and the server state deteriorates to the next state with each interruption and the previous server state is restored with each clearance action. At server state S_j, the interruptions arrive according to a Poisson process with rate f_j for $j = 1, \ldots, n-1$, and the clearance times are i.i.d. exponentials with rate r_j for $j = 2, \ldots, n$. Here $f_n = 0$ and $r_1 = 0$. The model considered above also includes the case that different types of failures might arrive at the normal state to transform the server state either to the moderate failure state or to the severe failure state depending on the severity of the incident. The clearance times of these incidents also depend on the incident type. Figure 3 presents this case where the server states are represented as N corresponding to the normal road conditions, M corresponding to the moderate incident and F corresponding to the severe incident conditions. The interruption and vehicle arrival processes, and the service and clearance times are all assumed to be mutually independent. In figures 2 and 3, $0 < i < C$ and $k \geq C$. Note that, for $C = 1$ the system considered here is a special case of the MAP/MSP/1 queue studied in [43], since in the later one the server state can go into any of the other server states.

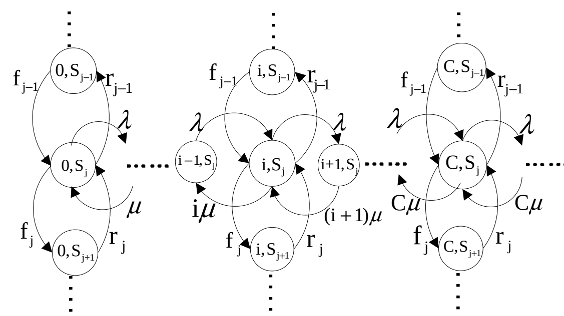

Figure 2: State transitions for M/M/C queue with deteriorating service

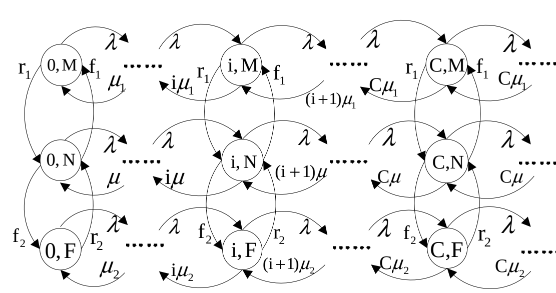

Figure 3: State transitions for M/M/C queue with three server states

The stochastic process $\{X(t), Y(t)\}$ describes the state of the link at time t, where $X(t)$ denotes the number of vehicles on the link at t, and $Y(t)$ denotes the server state.

Balance Equations

The steady-state balance equations are given below,
State S_1,

$$(\lambda + f_1)P_{0,S_1} = \mu_1 P_{1,S_1} + r_2 P_{0,S_2} \tag{1}$$

$$(\lambda + f_1 + i\mu_1)P_{i,S_1} = (i+1)\mu_1 P_{i+1,S_1} + r_2 P_{i,S_2} + \lambda P_{i-1,S_1} \quad (1 \le i \le C-1) \tag{2}$$

$$(\lambda + f_1 + C\mu_1)P_{i,S_1} = C\mu_1 P_{i+1,S_1} + r_2 P_{i,S_2} + \lambda P_{i-1,S_1} \quad (i \ge C) \tag{3}$$

State S_n,

$$(\lambda + r_n)P_{0,S_n} = \mu_n P_{1,S_n} + f_{n-1} P_{0,S_{n-1}} \tag{4}$$

$$(\lambda + r_n + i\mu_n)P_{i,S_n} = (i+1)\mu_n P_{i+1,S_n} + f_{n-1} P_{i,S_{n-1}} + \lambda P_{i-1,S_n} \quad (1 \le i \le C-1) \tag{5}$$

$$(\lambda + r_n + C\mu_n)P_{i,S_n} = C\mu_n P_{i+1,S_n} + f_{n-1} P_{i,S_{n-1}} + \lambda P_{i-1,S_n} \quad (i \ge C) \tag{6}$$

State S_j $(j = 2, ...n-1)$,

$$(\lambda + f_j + r_j)P_{0,S_j} = \mu_j P_{1,S_j} + r_{j+1} P_{0,S_{j+1}} + f_{j-1} P_{0,S_{j-1}} \tag{7}$$

$$(\lambda + f_j + r_j + i\mu_j)P_{i,S_j} = (i+1)\mu_j P_{i+1,S_j} + r_{j+1} P_{i,S_{j+1}} + f_{j-1} P_{i,S_{j-1}} + \lambda P_{i-1,S_j} \quad (1 \le i \le C-1) \tag{8}$$

$$(\lambda + f_j + r_j + C\mu_j)P_{i,S_j} = C\mu_j P_{i+1,S_j} + r_{j+1} P_{i,S_{j+1}} + f_{j-1} P_{i,S_{j-1}} + \lambda P_{i-1,S_j} \quad (i \ge C) \tag{9}$$

Generating Function

We will use the partial generating functions,

$$G_j(z) = \sum_{i=0}^{\infty} z^i P_{i,S_j},$$

to write the overall generating function as,

$$G(z) = \sum_{j=1}^{n} G_j(z).$$

By multiply the balance equations with z^i, and summing all equations for state S_j, we obtain,

$$[\lambda z(1-z) + f_1 z + C\mu_1(z-1)]G_1(z) - r_2 z G_2(z) = \sum_{i=0}^{C-1} (z-1)(C-i)\mu_1 P_{i,S_1} z^i, \tag{10}$$

$$[\lambda z(1-z) + r_n z + C\mu_n(z-1)]G_n(z) - f_{n-1} z G_{n-1}(z) = \sum_{i=0}^{C-1} (z-1)(C-i)\mu_n P_{i,S_n} z^i, \tag{11}$$

$$[\lambda z(1-z) + r_j z + f_j z + C\mu_j(z-1)]G_j(z) - r_{j+1} z G_{j+1}(z) - f_{j-1} z G_{j-1}(z) = \sum_{i=0}^{C-1} (z-1)(C-i)\mu_j P_{i,S_j} z^j, \quad (j = 2, 3, ...n-1) \tag{12}$$

In these n equations, there are nC unknown probabilities, and we can use the balance equations to reduce them to only n unknowns, P_{0,S_j}, for $j = 1, \dots, n$.

Proposition 1: For the n-state M/MSP/C queue, the stability condition is,

$$\lambda < \frac{\sum_{j=1}^{n} C\mu_j \cdot \left(\prod_{i=1}^{j-1} f_i \cdot \prod_{i=j+1}^{n} r_i\right)}{\sum_{k=1}^{n} \left(\prod_{i=1}^{k-1} f_i \cdot \prod_{i=k+1}^{n} r_i\right)}. \tag{13}$$

Proof: We know that $G_j(1)$ corresponds to the probability that the system is in server state S_j in the long run. If we aggregate all states (i, S_j) in server state S_j as a mega state, then we can easily obtain the long-run probability that the system is in state S_j as,

$$G_j(1) = \frac{\prod_{i=1}^{j-1} f_i \cdot \prod_{i=j+1}^{n} r_i}{\sum_{k=1}^{n} \left(\prod_{i=1}^{k-1} f_i \cdot \prod_{i=k+1}^{n} r_i\right)}. \tag{14}$$

Thus, the stability condition for this system is,

$$\lambda < \sum_{j=1}^{n} C\mu_j \cdot G_j(1), \tag{15}$$

giving the required inequality 13. □

In the next part, we will show that the denominator of $G(z)$ has $n-1$ distinct real roots that are unstable. These poles have to be eliminated by the zeros of $G(z)$, thus, giving $n-1$ equations in addition to $G(1) = 1$ to solve for the n unknowns. To this end, following the notation and the method introduced in [28], let,

$$g_1(z) = \lambda z(1-z) + f_1 z + C\mu_1(z-1),$$
$$g_j(z) = \lambda z(1-z) + r_j z + f_j z + C\mu_j(z-1), \quad (j = 2, 3, ...n-1),$$
$$g_n(z) = \lambda z(1-z) + r_n z + C\mu_n(z-1).$$

Further let,

$$A(z) = \begin{bmatrix} g_1(z) & -r_2 z & 0 & \cdots\cdots & 0 & 0 \\ -f_1 z & g_2(z) & -r_3 z & \cdots\cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots\cdots & -f_{n-1} z & g_n(z) \end{bmatrix}.$$

$$\vec{b}(z) = \begin{bmatrix} \sum_{i=0}^{C-1} (C-i)\mu_1 P_{i,S_1} z^i \\ \sum_{i=0}^{C-1} (C-i)\mu_2 P_{i,S_2} z^i \\ \vdots \\ \sum_{i=0}^{C-1} (C-i)\mu_n P_{i,S_n} z^i \end{bmatrix}, \quad \vec{G}(z) = \begin{bmatrix} G_1(z) \\ G_2(z) \\ \vdots \\ G_n(z) \end{bmatrix}.$$

Equations 10-12 can be written in the following compact form,

$$A(z)\vec{G}(z) = (z-1)\vec{b}(z).$$

It is easy to show that $A(z)$ has a singularity at $z = 1$. Since $|A(z)|$ is a polynomial (degree of $2n$) in z, we may write,

$$|A(z)| = (z-1)Q(z), \tag{16}$$

where $Q(z)$ is a polynomial of degree $2n-1$. Using Cramer's rule, for all values of z at which $A(z)$ is nonsingular, we have,

$$|A(z)|\, G_j(z) = |A_j(z)|(z-1), \qquad i = 1, 2, ...n. \tag{17}$$

Here, matrix $A_j(z)$ is obtained by replacing the jth column of $A(z)$ with $\vec{b}(z)$. The equation 17 must hold for all $z \in [0,1]$ since all functions in 17 are continuous and bounded in $[0,1]$, in addition the polynomial $|A(z)|$ may have only a finite number of roots in this interval.

The following lemma would be needed in the proof of Theorem 1.

Lemma 1: $Q(1) > 0$.

Proof: Using 16, equation 17 may be rewritten as,

$$Q(z)G_j(z) = |A_j(z)| \qquad j = 1, 2, ...n. \tag{18}$$

Taking the derivative of equation 16 with respect to z, then letting $z = 1$ gives,

$$Q(1) = \left.\frac{d|A(z)|}{dz}\right|_{z=1}. \tag{19}$$

Let $\vec{a}_j(z)$ be the jth row vector of matrix $A(z)$. We know that,

$$\left.\frac{d|A(z)|}{dz}\right|_{z=1} = \begin{vmatrix} \vec{a}_1'(1) \\ \vec{a}_2(1) \\ \vdots \\ \vec{a}_n(1) \end{vmatrix} + \begin{vmatrix} \vec{a}_1(1) \\ \vec{a}_2'(1) \\ \vdots \\ \vec{a}_n(1) \end{vmatrix} + \cdots\cdots + \begin{vmatrix} \vec{a}_1(1) \\ \vec{a}_2(1) \\ \vdots \\ \vec{a}_n'(1) \end{vmatrix}. \tag{20}$$

Using the definition of $A(z)$, we obtain,

$$\begin{vmatrix} \vec{a}_1(1) \\ \vdots \\ \vec{a}_j'(1) \\ \vdots \\ \vec{a}_n(1) \end{vmatrix} = (C\mu_j - \lambda) \cdot \prod_{i=1}^{j-1} f_i \prod_{i=j+1}^{n} r_i.$$

Then, from 19 and 20, we have,

$$Q(1) = \sum_{j=1}^{n} (C\mu_j - \lambda) \cdot \prod_{i=1}^{j-1} f_i \prod_{i=j+1}^{n} r_j. \tag{21}$$

The result follows from Proposition 1. □

From equations 16, 17, and the definition of G(z), clearly, the generating function of the number of customers in the system is,

$$G(z) = \frac{\sum_{j=1}^{n} |A_j(z)|}{Q(z)}. \tag{22}$$

Letting $z = 1$ in equation 18 gives,

$$|A_j(1)| = Q(1)G_j(1) \qquad j = 1, 2, ..., n. \tag{23}$$

$Q(1)$ and $G_j(1)$ are given by equations 21 and 14. The $n-1$ of the n equations in 23 are all redundant since multiplying $\frac{f_j}{r_{j+1}}$ to the jth equation of 23 will give the $(j+1)st$ equation. On the other hand, since $|A_j(z)|$ must be zero whenever $Q(z) = 0, 0 \le z < 1$, the next theorem proves that the generating function has $n-1$ unstable poles. Thus, the remaining equations will be obtained by equating the nominator of the generating function to zero at these unstable poles.

Theorem 1: The polynomial $Q(z)$ exactly has $n-1$ distinct real roots in the interval (0,1).

By Lemma 1, we have $Q(1) > 0$. Then, the proof follows from [28] since A matrix has a similar structure as the model considered in [28].

3. SPECIAL CASES

In this section we consider the case with a single failure state. Thus this case reduces to the queue with two-state Markov modulated service process considered in Baykal-Gursoy *et al.* [4]. Since there is only one faiure state we will use the failure and repair rates without any subscript as f and r. The service rate under normal conditions is denoted as μ and when the system failure occurs the service rate reduces to μ'. Also, let N denote the normal state, and F denote the failure state. Baykal-Gursoy *et al.*[4] obtained the generating function as,

$$G(z) = \frac{\begin{array}{c}[\lambda z(1-z) + C\mu'(z-1) + (r+f)z]\sum_{i+0}^{C-1} \mu z^i P_{i,N} \\ +[\lambda z(1-z) + C\mu(z-1) + (r+f)z]\sum_{i+0}^{C-1} \mu' z^i P_{i,F}\end{array}}{\begin{array}{c}\lambda^2 z^3 - (\lambda^2 + C\lambda\mu + \lambda f + C\lambda\mu' + \lambda r)z^2 \\ +(C\lambda\mu + C\lambda\mu' + C^2\mu\mu' + Cf\mu' + C\mu r)z - C^2\mu\mu'\end{array}}. \tag{24}$$

In this case, the stability condition is given as

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

By finding the roots of the denominator one of which is in-

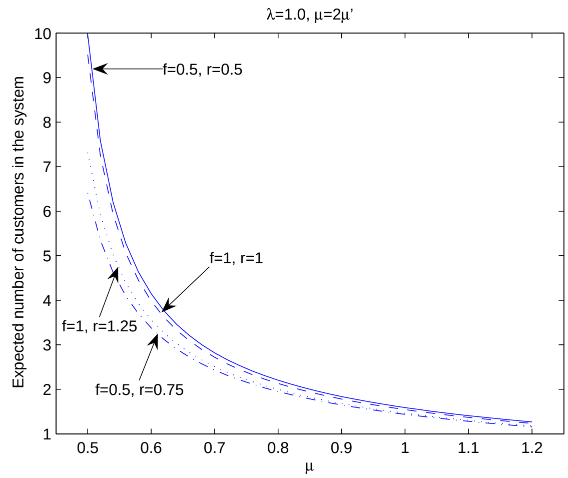

Figure 4: $\lambda = 1.0, \mu = 2\mu'$

side $(0, 1)$, we can obtain all of the unknown probabilities in the generating function. The expected number in the system is then obtained from $G'(1)$. Finding the single unstable root parametrically so that the generating function is obtained in closed form is elusive. Thus, in the case of partial failures $\mu' > 0$, this procedure is numerical. As an example, consider an M/M/3 queue subject to interruptions that reduce the service rate to a half of its normal value. The expected number of vehicles on the link versus the service rate μ is plotted in Figure 3. In this figure, $\lambda = 1.0, \mu 2\mu'$, and f and r take some particular values. It can be seen from Figure 3 that the number of vehicles on the link decreases as the service rate increases. Note that, for the two top most cases the stability condition requires that $\mu > 4/9\lambda$, since $r/(r+f) = 1/2$. If the service rate does not change, higher incident frequency or slower clearance rate would lead to more vehicles on the link. Figure 3 is used to show the effect of μ', where μ is fixed at 2 and μ' is increased independently. Similar to Figure 3, we let $\lambda = 1$, and f and r vary over a range. It can be seen that the expected number of customers also decrease as μ' increases, more significantly than in Figure 3, where μ and μ' increase simultaneously. Clearly, the stationary number of vehicles on the link when no incident occurs will constitute the lower bound.

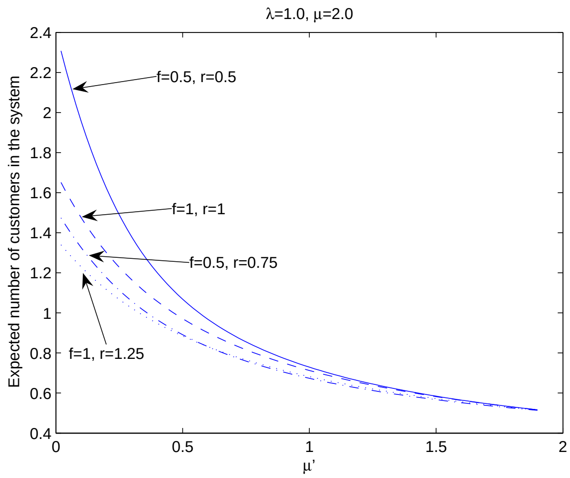

Figure 5: $\lambda = 1.0, \mu = 2.0$

On the other hand, closed form solutions can be obtained for the complete breakdown case as will be shown in the next part.

M/M/C Queues with System Breakdowns and Repairs ($\mu' = 0$)

As we have said before, the M/M/1 queue under complete server breakdown has been studied by Mitrani and Avi-Itzhak [28], and Gaver [13]. The generating function in this case can be written as,

$$G(z) = \frac{\frac{r}{r+f}(1-\rho\frac{r+f}{r})(1-\lambda/\delta z)}{(1-\rho z)(1-\lambda/\delta z)-\frac{f}{\delta}}, \tag{25}$$

where $\rho\frac{\lambda}{C\mu}$ with $C=1$ and $\delta = \lambda + r + f$. Since the generating function of regular M/M/1 queue without interruptions is $G(z) = \frac{1-\rho}{1-\rho z}$, we see from 25 that contrary to Doshi[8] this system does not exhibit the stochastic decomposition property.

For the M/M/2 queue, the generation function is given in (Baykal-Gursoy *et al.*[4]) as,

$$G(z) = \frac{\frac{r}{r+f}(1-\rho\frac{r+f}{r})(1-\lambda/\delta z)}{(1-\rho z)(1-\lambda/\delta z)-\frac{f}{\delta}} \cdot \frac{C+\eta z}{C+\eta}, \tag{26}$$

where $\eta = \frac{\lambda}{\mu}(\frac{\lambda+r+f}{\lambda+r})$.

Finally, we will consider the above M/M/3 queue with complete breakdowns. In this case, the generating function is,

$$G(z) = \frac{[\lambda(1-z)+r+f](3\mu P_{0,N}+2\mu z P_{1,N}+\mu z^2 P_{2,N})}{\lambda^2 z^2-(\lambda^2+\lambda r+f\lambda+3\lambda\mu)z+3\mu(\lambda+r)}. \tag{27}$$

Since $G(1) = 1$, equation 27 provides,

$$3P_{0,N}+2P_{1,N}+P_{2,N} = -\frac{\lambda}{\mu}+\frac{3r}{r+f}. \tag{28}$$

Using the balance equations, we evaluate,

$$\begin{aligned}
P_{0,F} &= \frac{f}{\lambda+r}P_{0,N};\\
P_{1,N} &= \frac{\lambda(\lambda+r+f)}{\mu(\lambda+r)}P_{0,N} = \eta P_{0,N};\\
P_{1,F} &= \frac{f}{\lambda+r}P_{1,N}+\frac{\lambda f}{(\lambda+r)^2}P_{0,N}\\
&= \frac{\lambda f(\lambda+r+f+\mu)}{\mu(\lambda+r)^2}P_{0,N};\\
P_{2,N} &= \frac{\lambda^2(\lambda+r+f)^2+f\mu\lambda^2}{2\mu^2(\lambda+r)^2}P_{0,N}\\
&= \left(\frac{1}{2}\eta^2+\frac{f\lambda^2}{2\mu(\lambda+r)^2}\right)P_{0,N}.
\end{aligned}$$

By substituting the above probabilities in equation 28, we obtain,

$$P_{0,N} = \frac{3\frac{r}{r+f}(1-\rho\frac{r+f}{r})}{3+2\eta+\frac{1}{2}\eta^2+\frac{f\lambda^2}{2\mu(\lambda+r)^2}}.$$

We also have,

$$3P_{0,N}+2zP_{1,N}+z^2P_{2,N} = \frac{3r}{r+f}\left(1-\rho\frac{r+f}{r}\right) \cdot \frac{3+2\eta z+(\frac{1}{2}\eta^2+\frac{f\lambda^2}{2\mu(\lambda+r)^2})z^2}{3+2\eta+\frac{1}{2}\eta^2+\frac{f\lambda^2}{2\mu(\lambda+r)^2}}.$$

Thus, the final form of generating function is,

$$G(z) = \frac{(1-\frac{\lambda}{\delta}z)\frac{r}{r+f}(1-\rho\frac{r+f}{r})}{(1-\rho z)(1-\frac{\lambda}{\delta}z)-\frac{f}{\delta}} \cdot \frac{3+2\eta z+(\frac{1}{2}\eta^2+\frac{f\lambda^2}{2\mu(\lambda+r)^2})z^2}{3+2\eta+\frac{1}{2}\eta^2+\frac{f\lambda^2}{2\mu(\lambda+r)^2}}. \tag{29}$$

As the number of servers increases, this system converges to an infinite server queue. Infinite server queues are more amenable to analysis even in the case of partial failures. It is shown in (Baykal-Gursoy and Xiao [3]), that the generating function has the following closed form,

$$G(z) = e^{(\lambda/\mu)(z-1)}\Psi(z), \tag{30}$$

where $\Psi(z)$ is the generating function of the mixture of two independent random variables. Depending on the value of μ' these two random variables are either in the form of generalized negative binomials (for the complete breakdown case) or Poissons with means distributed as truncated beta (for the partial failure case). Clearly, this system (30) exhibits the decomposition property.

4. CONCLUSIONS

The analysis of M/MSP/C queue with n server states presented in this paper clearly indicates that explicit solutions for the general case would be difficult to obtain. But, numerical methods as shown, could always be applied. For the special case of system breakdowns and repairs ($\mu' = 0$), the explicit solutions are obtained. Because breakdowns might happen during the service time of customers, the service completion time, i.e., dwell time on a link, will not remain exponential. So, the system we are solving could be considered as an M/G/C queue with a special service structure.

There is little known about M/G/C queues that the closed form solutions obtained in [4] and this paper will help to fill this gap.

5. REFERENCES

[1] I. Adan and V. Kulkarni. Single-server queue with Markov-dependent inter-arrival and service times. *Queueing Systems*, 45:113–134, 2003.

[2] B. Avi-Itzhak and P. Naor. Some queueing problems with the service station subject to breakdown. *Operations Research*, 11:303–320, 1963.

[3] M. Baykal-Gursoy and W. Xiao. Stochastic decomposition in M/M/∞ queues with Markov-modulated service rates. *Queueing Systems*, 48:75–88, 2004.

[4] M. Baykal-Gursoy, W. Xiao, and K. Ozbay. Modeling traffic flow interrupted by incidents, submitted for publication. I& SE-Working paper 05-024, Industrial and Systems Engineering Department, Rutgers University, 2005.

[5] O. Boxma and I. Kurkova. The M/M/1 queue in a heavy-tailed random environment. *Statistica Neerlandica*, 54(2), 2000.

[6] O. Boxma and I. Kurkova. The M/G/1 queue with two service speeds. *Advances in Applied Probability*, 33:520–540, 2001.

[7] J. Cheah and J. Smith. Generalized M/G/C/C state dependent queuing models and pedestrian traffic flows. *Queueing Systems*, 15:365–385, 1994.

[8] B. Doshi. Single server queues with vacations. In H. Takagi, editor, *Stochastic Analysis of Computer and Communication Systems*, pages 217–265. 1990.

[9] M. Eisen and M. Tainiter. Stochastic variations in queueing processes. *Operations Research*, 11:6:922–927, 1963.

[10] A. Federgruen and L. Green. Queueing systems with service interruptions. *Operations Research*, 34:752–768, 1986.

[11] A. Federgruen and L. Green. Queueing systems with service interruptions ii. *Naval Research Logistics*, 35:345–358, 1988.

[12] M. Fischer. An approximation to queueing systems with interruptions. *Management Science*, 24:338–344, 1977.

[13] D. Gaver. A waiting line with interrupted service, including priorities. *J. Roy. Stat. Soc. B24*, pages 73–90, 1962.

[14] D. Gazis, R. Herman, and R. Potts. Car-following theory of steady-state traffic flow. *Operations Research*, 7:499–505, 1959.

[15] D. Gazis, R. Herman, and R. Rothery. Nonlinear follow the leader models of traffic flow. *Operations Research*, 9:545–567, 1961.

[16] S. Halfin. Steady-state distribution for the buffer content of an M/G/1 queue with varying service rate. *SIAM J. Appl. Math.*, 23:356–363, 1972.

[17] D. Heidemann. A queueing theory approach to speed-flow-density relationships. In *Proc. Of the 13th International Symposium on Transportation and Traffic Theory*, France, July 1996.

[18] D. Heidemann. A queueing theory model of nonstationary traffic flow. *Transportation Science*, 35:405–412, 2001.

[19] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics 73*, pages 1067–1124, 2001.

[20] R. Jain and J. Smith. Modeling vehicular traffic flow using M/G/C/C state dependent queueing models. *Transportation Science*, 31:324–336, 1997.

[21] A. K. Jayawardene and O. Kella. M/G/∞ with alternating renewal breakdowns. *Queueing Systems*, 22:79–95, 1996.

[22] J. Keilson. Queues subject to service interruptions. *Ann. Math. Statistics*, 33:1314–1322, 1962.

[23] J. Keilson and L. Servi. The matrix M/M/∞ system: Retrial models and Markov modulated sources. *Advances in Applied Probability*, 25:453–471, 1993.

[24] M. Lighthill and G. Whitham. On kinematic waves: Ii. a theory of traffic on long crowded roads. In *Proc. Roy. Soc. London Ser. A 229*, pages 317–345, 1955.

[25] S. Mahabhashyam and N. Gautam. On queues with Markov-modulated service rates. *Queueing Systems*, 51:1-2:89–113, 2005.

[26] A. May and H. Keller. A deterministic queueing model. *Transp. Res.*, 1:2:117–128, 1967.

[27] A. May and H. Keller. Non-integer car-following models. *Highway Res. Rec.*, 199:19–32, 1967.

[28] I. Mitrani and B. Avi-Itzhak. A many-server queue with service interruptions. *Operations Research*, 16:628–638, 1968.

[29] NCTIM. In *National Conference on Traffic Incident Management: A Road Map to the Future*, pages 2–4, March 2002.

[30] M. Neuts. Further results on the M/M/1 queue with randomly varying rates. *OPSEARCH*, 15:4:158–168, 1978.

[31] M. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The John Hopkins University Press, 1981.

[32] M. Neuts. *Structured Stochastic Matrices of MG/G/1 Type and Their Applications*. Marcel Dekker, New York, 1989.

[33] G. Newell. Approximation methods for queues with application to the fixed-cycle traffic light. *SIAM Rev.*, 7:2:223–240, 1965.

[34] G. Newell. *Applications of Queueing Theory.* Chapman and Hall, London, 1971.

[35] C. O'Cinneide and P. Purdue. The M/M/∞ queue in a random environment. *J. Appl. Probab.*, 23:175–184, 1986.

[36] T. Ozawa. Analysis of queues with markovian service processes. *Stochastic Models*, 20:4:391–413, 2004.

[37] I. Prigogine and R. Herman. *Kinetic Theory of Vehicular Traffic*. Elsevier, NY, 1971.

[38] P. Purdue. The M/M/1 queue in a Markovian environment. *Operations Research*, 22:562–569, 1973.

[39] P. Richards. Shock waves on the highway. *Operations Research*, 4:42–51, 1956.

[40] A. Skabardonis, K. Petty, P. Varaiya, and R. Bertini. Evaluation of the Freeway Service Patrol (FSP) in Los Angeles, ucb-its-prr-98-31. Technical report, California PATH Research Report, Institute of Transportation Studies, University of California, Berkeley, 1998.

[41] N. Vandaele, T. VanWoensel, and N. Verbruggen. A queueing based traffic flow model. *Transportation Research-D: Transportation and Environment*, 5:121–135, 2000.

[42] H. White and L. Christie. Queuing with preemptive priorities or with breakdown. *Operations Research*, 6:79–95, 1958.

[43] U. Yechiali. A queueing-type birth-and-death process defined on a continuous-time Markov chain. *Operations Research*, 21:604–609, 1973.

[44] U. Yechiali and P. Naor. Queueing problems with heterogeneous arrivals and service. *Operations Research*, 19:722–734, 1971.