A Note on Infinite-Server Markov Modulated and Single-Server Retrial Queues

Zhe Duan and Melike Baykal-Gürsoy

Industrial and Systems Engineering Department, Rutgers University

Abstract

We reconsider the M/M/∞ queue with two-state Markov modulated arrival and service processes and the single-server retrial queue analyzed in Keilson and Servi [1]. Fuhrmann and Cooper type stochastic decomposition holds for the stationary occupancy distributions in both queues ([1] and [2]). The main contribution of the present paper is the derivation of the explicit form of the stationary system size distributions. Numerical examples are presented visually exhibiting the effect of various parameters on the stationary distributions.

Index Terms

Markov modulated, stochastic decomposition, Kummer functions, retrial queue

I. Introduction

We reconsider the article by Keilson and Servi [1]. They study a two-state Markov modulated arrival and service queue and a single-server retrial queue and obtain the generating functions for the stationary system size distributions. These types of systems are applicable to traffic problems [2], [3], [4], [5], service systems including call centers [6], [7], internet traffic [8], [9], [10], [11] and wireless systems [12], [13], [14].

In Markov modulated queuing systems, arrival and/or service processes are modulated by an external random environment process. Such a random environment, in fact, also models behavioral dependency of interarrival and service times in each state of the modulation process. For example, one may introduce a Poisson arrival process with two different arrival rates λ_1, λ_2 and an exponential service process with two different rates μ_1, μ_2. Depending on the random environment, load to the system is $\frac{\lambda_1}{\mu_1}$ or $\frac{\lambda_2}{\mu_2}$. Such queues may represent the traffic flow that is subjected to incidents. Incidents might happen randomly and may affect the whole system instantaneously, slowing every server. Meanwhile, throughout the incident clearance time, the arrival rate to the system might also decrease

Zhe Duan is a Ph.D. candidate in the Industrial and Systems Engineering department, Rutgers University E-mail: zheduan@eden.rutgers.edu

Dr. Baykal-Gürsoy is a faculty member in the Industrial and Systems Engineering Department, and affiliated with CAIT, RUTCOR, Rutgers University, 96 Frelinghuysen Rd, Piscataway, NJ 08854-801 E-mail: gursoy@rci.rutgers.edu

from its original level [2], [3], [4], [5]. The random environment models are also used in the case of customer abandonment [6], [15], [16] and in processor sharing queues [17], [18].

A simple retrial queue, on the other hand, is a single-server system in which customers upon finding the server busy join the orbit to retry entering the system after a random time [1]. Retrial queues are used in modeling internet traffic and wireless network systems [8], [9], [10], [11], [12], [13], [14]. For a summary of previous research on retrial queues, please refer to Yang [19], Falin [20], Falin and Templeton [21], Artalejo and Gómez-Corral [22] and Artalejo [23].

In this paper, we present full description of the stationary distributions of the system size for a Markov-modulated arrival and service (two-state system) M/M/∞ queue, and a single-server retrial queue with customer abandonments. Also, numerical examples are provided to explain visually how the behavior of the distribution changes with changing parameter values. This paper is organized as follows. Next, we briefly review the prior work. In Section 2, we consider a M/M/∞ queue in a two-state random environment. In Section 3, we present our results for the retrial queue. Finally, in Section 4, we discuss our results and future research directions.

A. Literature Review

Extensive research has been done on queueing systems with Markov modulated arrival process (MAP) (e.g. Takahashi and Wang [24], Ahn and Jeon [25], Kim [26]). Recently, queueing systems with Markovian Service process (MSP) have been receiving considerable attention, mainly due to their applicability to telecommunication and transportation systems. The articles on MSP queueing systems include Baykal-Gürsoy and Xiao [2], Boxma and Kurkova [18], Mahabhashyam and Gautam [27], Bekker and Boxma [28], [29]. Baykal-Gürsoy and Xiao [2] analyze a M/M/∞ queue with Markov modulated service rates, and its application to traffic modeling is introduced. In [18], a M/G/1 queue with two alternating service speeds is studied, explicit tail asymptotics for the tail of the buffer content distribution are obtained under some regularity conditions. Mahabhashyam and Gautam [27] present two interesting examples on queues with MSP systems, a Web-server with multi-class requests and a CPU with multiple processes. Bekker and Boxma [28], [29] consider queueing systems with adaptable service speed that depends on the system workload or the number of customers.

There are also articles that study varying arrival and service rates concurrently, such as Takine [30] and Bekker [31]. Takine [30] studies a queueing system with MAP and MSP, but the arrival process is a special MAP, called marked MAP. For this queueing system, a new queue with constant service speed is constructed by rescaling time, and the waiting time distribution for the original queue is derived. In [31], a M/G/1 queue with workload-dependent service and arrival is analyzed, and steady-state workload density is obtained.

There is a vast literature on retrial queues due to their applicability to various areas from telecommunication, supply chain, call centers, and other service systems. In this paper, we will concentrate on single server queues.

It has been known that some queuing systems possess the so-called stochastic decomposition property, i.e., it has the Fuhrmann-Cooper type representation. This property implies that a performance measure can be decomposed into much simpler and independent random variables. Fuhrmann and Cooper [32] show that the decomposing property holds for a M/G/1 queue with generalized vacations. The stationary number of customers is distributed as the sum of two independent random variables, one of which is the stationary number of customers in the standard M/G/1 queue. They also show the decomposition property for the stationary waiting times in a FIFO discipline vacation system. Later on such decomposition results have been shown for more general vacation systems [33], [34], [35], for M/G/1 retrial queues [19], [36], [37], [38], for single-server retrial queues with an unreliable server [39], [40], [41], [42], [43], [44].

Keilson and Servi [1] study a matrix MAP/MSP/1 and a retrial queue with customer abandonments. For the special case of a MAP/M/1 queue with a two-state Markov modulated arrival process, they show that the decomposition property holds, and provide the explicit solution. A M/G/∞ queue with alternating renewal breakdowns is studied in Jayawardene and Kella [45]; who show that the decomposition property holds.

Even though stochastic decomposition results have been established for certain queueing models, concise description of the component random variables may be needed. Baykal-Gürsoy and Xiao [2] consider a M/M/∞ system with a two state Markov-modulated service process, e.g., a M/MSP/∞ queue. Using the method introduced in [1], they show that this model also exhibits the stochastic decomposition property, and express the stationary distribution of the queue size in closed form. Recently, other decomposition results have been established for more general infinite server queues operating in random environment [6], [46], [47], [48], [49].

II. Markov-Modulated Arrival and Service

In this section, we consider a MAP/MSP/∞ queueing system studied in Keilson and Servi [1]. The arrival process is a Markov-modulated Poisson process, and the service process for all servers is also Markov modulated. The underlying stochastic process can be modeled as $\{X(t), U(t)\}$ where $X(t)$ is the number of customers in the system and $U(t)$ is the functioning state of the system at time t as shown in Fig. 1.

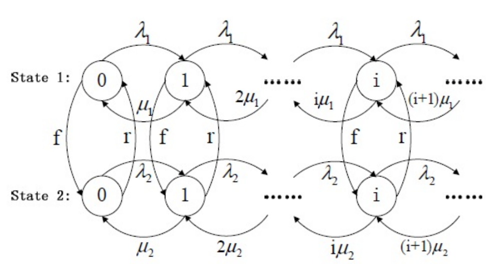

Fig. 1. Rate diagram of the Markov-modulated M/M/∞ queue

State 1 represents the system when it functions normally. When an interruption occurs, system transits to state 2. The arrival of customers is a Poisson process with rate λ_1 (λ_2) and service times are independent and identically distributed (i.i.d.) exponentially with rate μ_1 (μ_2) when the system state is 1 (2). The time to interruption is distributed exponentially with rate f. The time to switch back to normal state is also distributed exponentially with rate r. All transition times are independent of each other.

We assume that $\mu_1 > 0$ and $\frac{\lambda_1}{\mu_1} < \frac{\lambda_2}{\mu_2}$ without loss of generality. We will analyze two cases; $\mu_2 > 0$, and $\mu_2 = 0$. For the first case, i.e., $\mu_2 > 0$, we have

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}; \qquad \Gamma = \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix}; \qquad Q = \begin{bmatrix} -f & f \\ r & -r \end{bmatrix}. \tag{1}$$

Keilson and Servi [1] obtain the probability generating function of the number in the system as,

$$\begin{aligned}\pi(u) = exp\left[\tfrac{\lambda_1}{\mu_1}(u-1)\right] \cdot \Bigg[& \frac{f\mu_2 + r\mu_1}{\mu_2(f+r)} \cdot M\left(\frac{f}{\mu_1},\ \frac{f}{\mu_1} + \frac{r}{\mu_2},\ \left(\frac{\lambda_2}{\mu_2} - \frac{\lambda_1}{\mu_1}\right)(u-1)\right) \\ & + \frac{r(\mu_2 - \mu_1)}{\mu_2(f+r)} \cdot M\left(\frac{f}{\mu_1},\ \frac{f}{\mu_1} + \frac{r}{\mu_2} + 1,\ \left(\frac{\lambda_2}{\mu_2} - \frac{\lambda_1}{\mu_1}\right)(u-1)\right)\Bigg].\end{aligned} \tag{2}$$

Using equation 13.4.3 in [50], [51], we derive another form of this probability generating function,

$$\begin{aligned}\pi(u) = exp\left[\tfrac{\lambda_1}{\mu_1}(u-1)\right] \cdot \Bigg[& \frac{f\mu_2 + r\mu_1}{\mu_1(f+r)} \cdot M\left(\frac{f}{\mu_1},\ \frac{f}{\mu_1} + \frac{r}{\mu_2},\ \left(\frac{\lambda_2}{\mu_2} - \frac{\lambda_1}{\mu_1}\right)(u-1)\right) \\ & + \frac{f(\mu_1 - \mu_2)}{\mu_1(f+r)} \cdot M\left(\frac{f}{\mu_1} + 1,\ \frac{f}{\mu_1} + \frac{r}{\mu_2} + 1,\ \left(\frac{\lambda_2}{\mu_2} - \frac{\lambda_1}{\mu_1}\right)(u-1)\right)\Bigg].\end{aligned} \tag{3}$$

Theorem 1 *The stationary number of customers in the system, X, for ($\mu_1, \mu_2 > 0$ and $\frac{\lambda_1}{\mu_1} < \frac{\lambda_2}{\mu_2}$) has the form*

$$X = X_\varphi + Y, \tag{4}$$

where X_φ and Y are independent. X_φ is the stationary number of customers in an M/M/∞ queue, thus, is Poisson distributed with mean $\varphi = \lambda_1/\mu_1$. Y is the additional accumulation of customers due to varying arrival and service processes. It is given as

$$P\{Y = n\} = pP\{Y_1 = n\} + (1-p)P\{Y_2 = n\}. \tag{5}$$

(i) If $\mu_1 < \mu_2$, $p = (f\mu_2 + r\mu_1)/(\mu_2(f+r))$, and Y_1 and Y_2 are conditionally Poisson distributed with random means that have truncated beta distributions $B(a, b, -2\rho*)$ and $B(a, b+1, -2\rho*)$, respectively, where,

$$a = \frac{f}{\mu_1}, \qquad b = \frac{f}{\mu_1} + \frac{r}{\mu_2}, \qquad \rho* = \frac{1}{2}\left(\frac{\lambda_1}{\mu_1} - \frac{\lambda_2}{\mu_2}\right). \tag{6}$$

(ii) When $\mu_1 > \mu_2$, $p = (f\mu_2 + r\mu_1)/(\mu_1(f+r))$, and Y_1 and Y_2 are conditionally Poisson distributed with random means that have truncated beta distributions $B(a, b, -2\rho*)$ and $B(a+1, b+1, -2\rho*)$, respectively. a, b and $\rho*$ are the same as in equation (6).

For details of the probability mass functions of random variables, Y_1 and Y_2, please see Baykal-Gürsoy and Xiao [2].

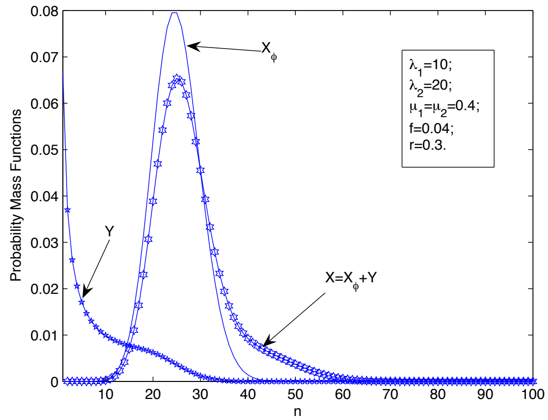

Fig. 2. Probability mass function of the stationary number of customers in the system (MAP)

Proof: Noticing that *Kummer* function, $M(a, b, c(u-1))$, is the generating function of a Poisson random variable randomized by truncated beta $B(a, b, c)$ [52], part (i) follows immediately from Eq. 2, and part (ii) from Eq. 3 (also see [2]). $\square$

We plot the stationary distributions for various cases. These examples are provided to explain visually how the behavior of the distribution changes with changing parameter values. Figure 2 shows the effect of Markov modulated arrivals, while figure 3 shows the combined effect of MAP and MSP. Both cause thickening at the tail of the distribution compared to the Poisson case.

The expectation and variance of the stationary number of customers in the system are as follows.

Corollary 1 *The expectation of the stationary number of customers in the system, X, for ($\mu_1, \mu_2 > 0$ and $\frac{\lambda_1}{\mu_1} < \frac{\lambda_2}{\mu_2}$),*

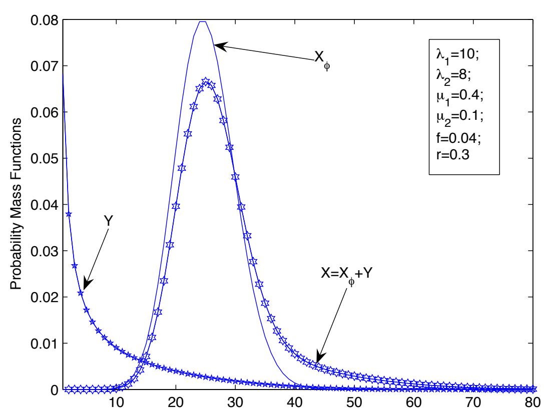

Fig. 3. Probability mass function of the stationary number of customers in the system (General Case: MAP/MSP)

is given as:

$$E(X) = \frac{\lambda_1}{\mu_1} + \frac{(\lambda_2\mu_1 - \lambda_1\mu_2)f(f+r+\mu_1)}{\mu_1(f+r)(f\mu_2 + r\mu_1 + \mu_1\mu_2)}, \tag{7}$$

and its variance is derived as:

$$\begin{aligned} Var(X) = & \frac{\lambda_1}{\mu_1} + \frac{f(\lambda_2\mu_1 - \lambda_1\mu_2)(f+r+\mu_1)}{\mu_1(f+r)(f\mu_2 + r\mu_1 + \mu_1\mu_2)} \\ & + \frac{fr(\lambda_2\mu_1 - \lambda_1\mu_2)^2 \cdot [(f+r)^2 + 2(f\mu_2 + r\mu_1 + \mu_1\mu_2) + f\mu_1 + r\mu_2]}{(f+r)^2(f\mu_2 + r\mu_1 + \mu_1\mu_2)^2 \cdot (f\mu_2 + r\mu_1 + 2\mu_1\mu_2)}. \end{aligned} \tag{8}$$

Proof: By taking the first and second derivatives of the generating function, and utilizing the formulas in [50], [51], and evaluating at $u = 1$, we can express 1st and 2nd factorial moments of X. The expectation and variance are then obtained using the factorial moments. □

We have the following two special cases:

CASE 1 - Markov Modulated Arrival only ($\mu_1 = \mu_2 = \mu$): (This case was considered in [1]) Then, since $\lambda_1/\mu_1 < \lambda_2/\mu_2$, we have $\lambda_1 < \lambda_2$. The generating function for the number of customers in the system is,

$$\pi(u) = exp\left[\frac{\lambda_1}{\mu}(u-1)\right] \cdot M\left(\frac{f}{\mu},\ \frac{f+r}{\mu},\ \frac{\lambda_2 - \lambda_1}{\mu}(u-1)\right). \tag{9}$$

The stochastic decomposition result is immediate. The first term is the generating function of a Poisson random variable with mean $\varphi = \lambda_1/\mu$, i..e., X_φ. We have the following stationary distribution for the additional customer accumulation. Y is conditionally Poisson distributed with random means that have truncated beta distributions $B(a, b, -2\rho*)$, with

$$a = \frac{f}{\mu}, \qquad b = \frac{f+r}{\mu}, \qquad \rho* = \frac{\lambda_1 - \lambda_2}{2\mu}. \tag{10}$$

CASE 2 - Markov Modulated Service only ($\lambda_1 = \lambda_2 = \lambda$): (This case was considered in [2]) Then, since $\lambda_1/\mu_1 < \lambda_2/\mu_2$, we have $\mu_1 > \mu_2$. The generating function for the number of customers in the system will be, (since $\mu_1 > \mu_2$, we use equation 3

$$\begin{aligned} \pi(u) = exp\left[\tfrac{\lambda}{\mu_1}(u-1)\right] \cdot \Bigg[& \frac{f\mu_2 + r\mu_1}{\mu_1(f+r)} \cdot M\left(\frac{f}{\mu_1},\ \frac{f}{\mu_1} + \frac{r}{\mu_2},\ \left(\frac{\lambda}{\mu_2} - \frac{\lambda}{\mu_1}\right)(u-1)\right) \\ & + \frac{r(\mu_1 - \mu_2)}{\mu_1(f+r)} \cdot M\left(\frac{f}{\mu_1} + 1,\ \frac{f}{\mu_1} + \frac{r}{\mu_2} + 1,\ \left(\frac{\lambda}{\mu_2} - \frac{\lambda}{\mu_1}\right)(u-1)\right)\Bigg]. \end{aligned} \tag{11}$$

The stochastic decomposition result together with the complete stationary distribution of the queue size are given in Baykal-Gursoy and Xiao [2].

The set of figures from 4 to 6, show that increasing r diminishes the impact of low service periods on the stationary distribution. That is because the probability that the system is in normal state, i.e., state 1, $\frac{r}{f+r}$, increases as r increases.

The case in which $\mu_2 = 0$, i.e., the singular case, is also studied in Keilson and Servi [1]. The probability

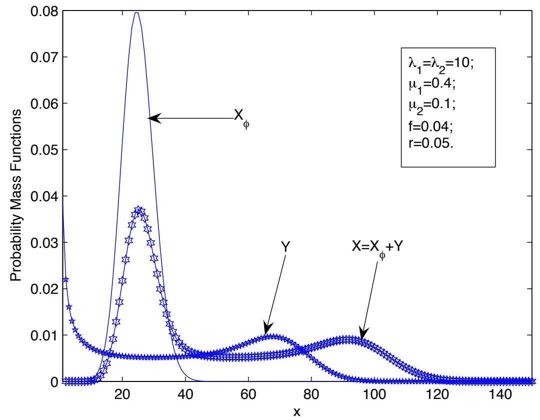

Fig. 4. Probability mass function of the stationary number of customers in the system (MSP, r=0.05)

generating function for the stationary queue size is obtained as [1],

$$\pi(u) = C_1 \cdot e^{\frac{\lambda_1}{\mu_1}u} \cdot [-\lambda_2\mu_1 u + (f + r + \lambda_2)\mu_1] \cdot [-\lambda_2\mu_1 u + \lambda_2\mu_1 + r\mu_1]^{-\frac{\mu_1+f}{\mu_1}}. \tag{12}$$

Theorem 2 *The stationary number of customers in the system, X, for ($\mu_1 > 0, \mu_2 = 0$ and $\frac{\lambda_1}{\mu_1} < \frac{\lambda_2}{\mu_2}$), can be decomposed as*

$$X = X_\varphi + Y, \tag{13}$$

where X_φ and Y are independent, X_φ is a Poisson random variable with mean $\varphi = \lambda_1/\mu_1$, and

$$P\{Y = n\} = pP\{Y_1 = n\} + (1-p)P\{Y_2 = n\}, \tag{14}$$

where $p = r/(r+f)$, Y_1 and Y_2 are generalized negative binomial or Polya distributed with $NB(f/\mu_1, \lambda_2/(\lambda_2+r))$ and $NB(f/\mu_1 + 1, \lambda_2/(\lambda_2 + r))$, respectively.

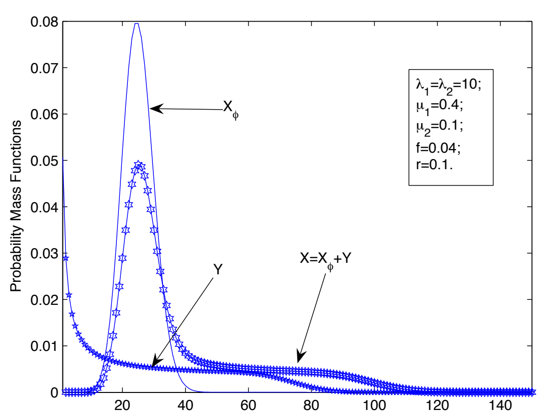

Fig. 5. Probability mass function of the number of customers in the system (MSP, r=0.1)

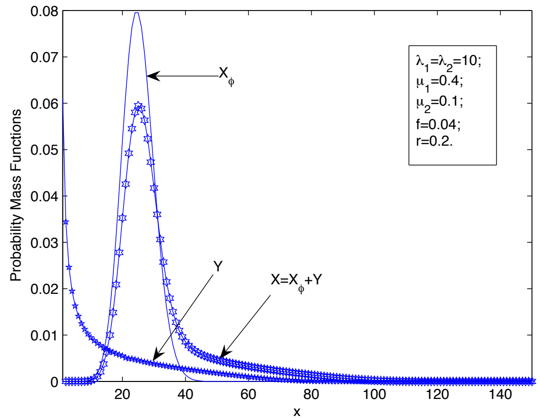

Fig. 6. Probability mass function of the stationary number of customers in the system (MSP, r=0.2)

The expected value and the variance of the stationary queue size are given as follows:

$$E[X] = \frac{1}{\mu_1}(\lambda_1 + \lambda_2 \frac{f}{r}(1 + \frac{\mu_1}{r+f})), \tag{15}$$

$$Var[X] = \frac{1}{\mu_1}(\lambda_1 + \lambda_2 \frac{f}{r}\frac{\lambda_2 + r}{r}(1 + \frac{\mu_1}{r+f})). \tag{16}$$

Proof: Using $\pi(1) = 1$, the value of constant C_1 is obtained as

$$C_1 = \frac{e^{-\frac{\lambda_1}{\mu_1}}(r\mu_1)^{\frac{\mu_1+f}{\mu_1}}}{(f+r)\mu_1}.$$

Then, we can simplify Eq.12 as

$$\pi(u) = e^{\frac{\lambda_1}{\mu_1}(u-1)} \left\{ \frac{r}{r+f} \left[\frac{r/(\lambda_2+r)}{1-\lambda_2 u/(\lambda_2+r)} \right]^{\frac{f}{\mu_1}} + \frac{f}{r+f} \left[\frac{r/(\lambda_2+r)}{1-\lambda_2 u/(\lambda_2+r)} \right]^{\frac{f}{\mu_1}+1} \right\}. \tag{17}$$

Noting that $\left[\frac{1-\alpha}{1-\alpha u}\right]^{\beta}$ is the probability generating function of the negative binomial random variable with parameters

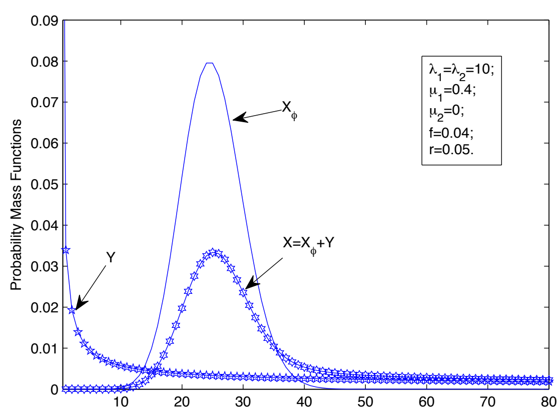

Fig. 7. Probability mass function of the stationary number of customers in the system (MSP, r=0.05)

α and β, the first part of the theorem is immediate. The expected value and the variance formulas also follow from this observation. Note that β here is a real number. $\square$

Figure 7 shows the adverse effect of service cessation on the tail of the queue size distribution.

III. Retrial Model

In this section, we study a M/M/1 retrial queue with no buffer space. The arrival process is Poisson with rate λ, and service times are i.i.d. exponential with rate μ. A customer upon arrival joins the orbit unless the server is idle. The time to retrial is exponentially distributed with rate ζ, and each customer will abandon the orbit after an exponential time with rate $\theta > 0$. All processes are assumed to be independent. This case is considered in Keilson and Servi [1], and the probability generating function of the number in the orbit is derived as

$$\pi(u) = K\left[M(b-a,b,-d-e(u-1)) + \frac{\lambda\zeta}{\lambda\theta+\mu(\theta+\zeta)}M(b-a+1,b+1,-d-e(u-a))\right], \tag{18}$$

where,

$$a = \frac{\mu}{\theta}, \qquad b = \frac{\lambda}{\theta+\zeta} + \frac{\mu}{\theta}, \qquad d = -\frac{\lambda\zeta}{\theta(\theta+\zeta)}, \qquad e = -\frac{\lambda}{\theta}.$$

We have the next theorem.

Theorem 3 *The stationary number of customers in the retrial orbit, Y, has the probability mass function as,*

$$P\{Y=n\} = p_1 \cdot P\{Y_1 = n\} + p_2 \cdot P\{Y_2 = n\}, \text{ for } \theta > 0, \tag{19}$$

where Y_1 and Y_2 are hyper-Poisson random variables with four parameters $(b-a, b, -d, -e)$ and $(b-a+1, b+1, -d, -e)$, respectively. p_1 and p_2 are given in equations 22 and 23.

Proof: Since $\pi(1) = 1$, we have

$$K = \frac{\lambda\theta+\mu(\theta+\zeta)}{[\lambda\theta+\mu(\theta+\zeta)]M(b-a,b,-d) + \lambda\zeta M(b-a+1,b+1,-d)}. \tag{20}$$

We rewrite the generating function, $\pi(u)$, as the convex combination of two partial generating functions

$$\pi(u) = p_1 \cdot \frac{M(b-a,b,-d-e(u-1))}{M(b-a,b,-d)} + p_2 \cdot \frac{M(b-a+1,b+1,-d-e(u-1))}{M(b-a+1,b+1,-d)}. \tag{21}$$

The coefficients, p_1 and p_2 are given by

$$p_1 = \frac{[\lambda\theta+\mu(\theta+\zeta)]M(b-a,b,-d)}{[\lambda\theta+\mu(\theta+\zeta)]M(b-a,b,-d) + \lambda\zeta M(b-a+1,b+1,-d)}, \tag{22}$$

$$p_2 = \frac{\lambda\zeta M(b-a+1,b+1,-d)}{[\lambda\theta+\mu(\theta+\zeta)]M(b-a,b,-d) + \lambda\zeta M(b-a+1,b+1,-d)}, \tag{23}$$

where $b - a = \frac{\lambda}{\theta+\zeta} > 0$ and $-d = \frac{\lambda\zeta}{\theta(\theta+\zeta)} > 0$.

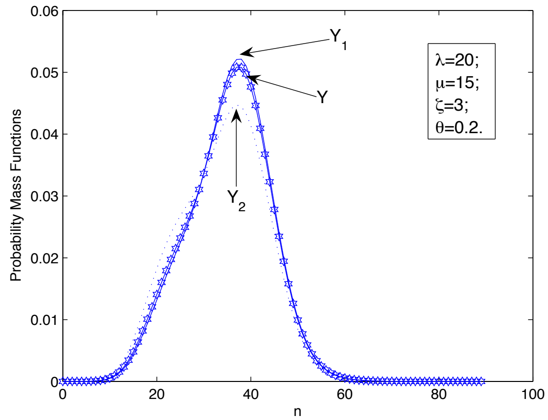

Fig. 8. Probability mass function of the number of customers in the orbit ($\zeta = 3$)

Let us denote the partial generating functions as $g_1(u)$ and $g_2(u)$, i.e.,

$$g_1(u) = \frac{M(b-a, b, -d-e(u-1))}{M(b-a, b, -d)}, \qquad g_2(u) = \frac{M(b-a+1, b+1, -d-e(u-1))}{M(b-a+1, b+1, -d)}.$$

We then recognize $g_1(u)$ and $g_2(u)$ as the generating functions of hyper-Poisson random variables with four parameters $(b-a, b, -d, -e)$ and $(b-a+1, b+1, -d, -e)$, respectively [52]. □

Note that the probability mass function of a hyper-Poisson random variable, N, with four parameters (a, b, c, λ) is

$$P\{N = n\} = \frac{M(a+n, b+n, c-\lambda)}{M(a, b, c)} \frac{(a)_n}{(b)_n} \frac{\lambda^n}{n!}. \tag{24}$$

Figures 8 to 10 represent the effect of increasing retrial rate, ζ, on the stationary distribution of Y. Although initially it is closer to the distribution of Y_1, as the retrial rate increases, it gets closer to the distribution of Y_2. Some special cases are as follows:

CASE 1 - $(b - a = b)$: If the first two parameters of an hyper-Poisson random variable are equivalent, the hyper-Poisson will recover the Poisson distribution. For example, when $b-a = b$, meaning $\mu = 0$, i.e., the server is down, customers leave the system through abandonment only. Hence, the stationary number in the orbit should follow Poisson distribution with rate $\frac{\lambda}{\theta}$. This can be seen by noticing that in this case $M(b-a, b, x) = e^x$, and,

$$\begin{aligned}
p_1 &= \frac{\lambda\theta \cdot e^{-d}}{\lambda\theta \cdot e^{-d} + \lambda\zeta \cdot e^{-d}} = \frac{\theta}{\theta + \zeta} > 0, \\
p_2 &= \frac{\zeta}{\theta + \zeta}, \\
g_1(u) &= \frac{e^{-d-e(u-1)}}{e^{-d}} = e^{\frac{\lambda}{\theta}(u-1)}, \\
g_2(u) &= e^{\frac{\lambda}{\theta}(u-1)},
\end{aligned}$$

giving,

$$\pi(u) = e^{\frac{\lambda}{\theta}(u-1)}. \tag{25}$$

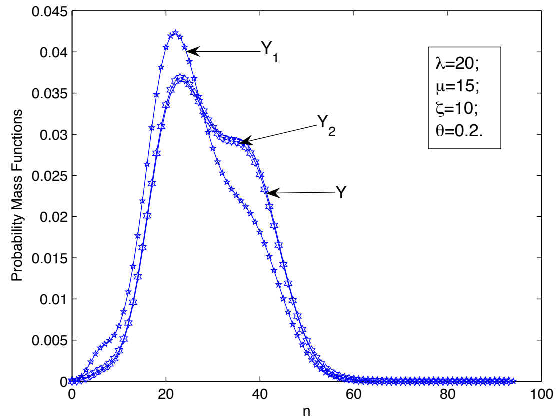

Fig. 9. Probability mass function of the stationary number of customers in the orbit ($\zeta = 10$)

CASE 2 - ($d = 0$): If $d = 0$, meaning $\zeta = 0$, then, $p_1 = 1, \quad p_2 = 0$. Thus,

$$\pi(u) = g_1(u) = M(b - a, b, -e(u - 1)),$$

is the generating function of a Poisson random variable randomized by truncated beta $B(b - a, b, -e)$.

CASE 3 - ($\theta = 0$): This case has been studied by Keilson *et al.* [53], and Jonin and Sodel [54]. The stationary distribution of the number of customers in the orbit has been described fully as the mixture of two negative binomial random variables.

Figures 8 to 10 represent the effect of increasing retrial rate, ζ, on the stationary distribution of Y.

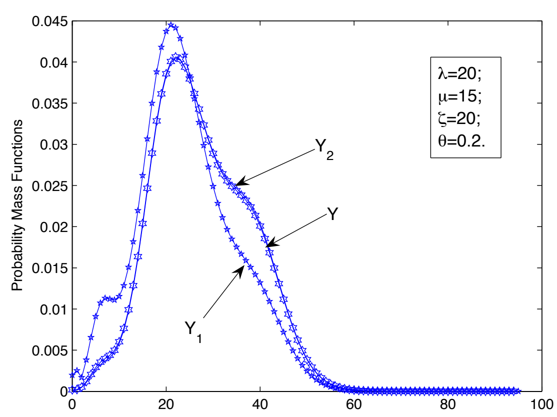

Fig. 10. Probability mass function of the stationary number of customers in the orbit ($\zeta = 20$)

IV. Discussion and future research

In this paper, we present the complete description of the stationary distribution of the queue size for M/M/∞ queue with two-state Markov modulated arrival and service processes. We derive its mean and variance. The stationary distribution of number in the orbit in a single-server retrial queue is also fully characterized. Special cases are discussed describing the effects of various parameters on the distributions.

Currently, we are studying infinite server queues in a more general random environment and infinite server retrial queues with deteriorating service. The first case corresponds to MAP/MSP/∞ with multiple environment states.

Acknowledments

The authors would like to thank the anonymous referee for insightful comments, suggestions and corrections.

References

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

[2] M. Baykal-Gürsoy and W. Xiao, "Stochastic decomposition in M/M/∞ queues with Markov-modulated service rates," *Queueing Systems*, vol. 48, pp. 75–88, 2004.

[3] M. Baykal-Gürsoy and Z. Duan, "M/M/C queues with Markov modulated service processes," in *First International Conference on Performance Evaluation Methodologies and Tools-Valuetools, Pisa, IT*, 2006.

[4] M. Baykal-Gürsoy, W. Xiao, and K. M. A. Ozbay, "Modeling traffic flow interrupted by incidents," *European Journal of Operational Research*, vol. 195, pp. 127–138, 2009.

[5] M. Baykal-Gürsoy, Z. Duan, and H. Xu, "Stochastic models of traffic flow interrupted by incidents," in *IFAC Symposium on Transportation Systems*, 2009.

[6] G. Pang and W. Whitt, "Heavy-traffic limits for many-server queues with service interruptions," *Queueing Systems*, vol. 61, no. 2, pp. 167–202, 2009.

[7] ——, "Service interruptions in large-scale service systems," *Management Science*, vol. 55, no. 9, pp. 1499–1512, 2009.

[8] J. R. Artalejo, D. S. Orlovsky, and A. N. Dudin, "Multi-server retrial model with variable number of active servers," *Computers & Industrial Engineering*, vol. 48, no. 2, pp. 273–288, 2005.

[9] K. Avrachenkov and U. Yechiali, "Retrial networks with finite buffers and their application to internet data traffic," *Probability in the Engineering and Informational Sciences*, vol. 22, no. 04, pp. 519–536, 2008.

[10] ——, "On tandem blocking queues with a common retrial queue," *Computers & Operations Research*, vol. 37, no. 7, pp. 1174–1180, 2010.

[11] C. Kim and S. Dudin, "A multi-server queueing model with retrial connection arrivals as a model for optimization of the traffic control," *International Journal of Systems Science*, vol. 43, no. 8, pp. 1555–1567, 2012.

[12] I. Atencia and P. Moreno, "A single-server retrial queue with general retrial times and Bernoulli schedule," *Applied Mathematics and Computation*, vol. 162, no. 2, pp. 855–880, 2005.

[13] P. Wüchner, J. Sztrik, and H. De Meer, "Finite-source M/M/S retrial queue with search for balking and impatient customers from the orbit," *Computer Networks*, vol. 53, no. 8, pp. 1264–1273, 2009.

[14] J. R. Artalejo and M. Lopez-Herrero, "Cellular mobile networks with repeated calls operating in random environment," *Computers & Operations research*, vol. 37, no. 7, pp. 1158–1166, 2010.

[15] U. Yechiali, "Queues with system disasters and impatient customers when system is down," *Queueing Systems*, vol. 56:3/4, pp. 195–202, 2007.

[16] N. Perel and U. Yechiali, "Queues with slow servers and impatient customers," *European Journal of Operational Research*, vol. 201, no. 1, pp. 247–258, 2010.

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

[18] ——, "The M/G/1 queue with two service speeds," *Advances in Applied Probability*, vol. 33, pp. 520–540, 2001.

[19] T. Yang and J. Templeton, "A survey on retrial queues," *Queueing Systems*, vol. 2, pp. 201–233, 1987.

[20] G. I. Falin, "A survey of retrial queues," *Queueing Systems*, vol. 7, pp. 127–168, 1990.

[21] G. I. Falin and J. G. C. Templeton, *Retrial Queues*. Chapman & Hall/CRC, 1997, vol. 75.

[22] J. R. Artalejo and A. Gómez-Corral, *Retrial Queueing Systems: A Computational Approach*. Springer Verlag, 2008.

[23] J. R. Artalejo, "Accessible bibliography on retrial queues: Progress in 2000-2009," *Mathematical and Computer Modelling*, vol. 51, no. 9-10, pp. 1071–1081, 2010.

[24] H. Takahashi and L. Wang, "Approximate analysis of a queueing system with Markov modulated arrivals," *Electronics and Communications in Japan, Part 1*, vol. 73, pp. 12–21, 1990.

[25] S. Ahn and J. Jeon, "Large deviation results for queues with continuous time Markov modulated arrivals," *Engineering Simulation*, vol. 18, pp. 67–74, 2000.

[26] B. Kim, J. Lee, and I. S. Wee, "Tail asymptotics for the fundamental period in the map/g/1 queue," *Queueing Systems*, vol. 57, no. 1, pp. 1–18, 2007.

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

[28] R. Bekker, O. J. Boxma, and J. Resing, "Queues with service speed adaptations," EURANDOM, Tech. Rep., 2006. [Online]. Available: http://www.eurandom.tue.nl/reports/2006/028OBreport.pdf

[29] R. Bekker and O. J. Boxma, "An M/G/1 queue with adaptable service speed," *Stochastic Models*, vol. 23, pp. 373–396, 2007.

[30] T. Takine, "Single-server queues with Markov-modulated arrivals and service speed," *Queueing Systems*, vol. 49, pp. 7–22, 2005.

[31] R. Bekker, "Finite-buffer queues with workload-dependent service and arrival rates," *Queueing Systems*, vol. 50, pp. 231–253, 2005.

[32] S. Fuhrmann and R. Cooper, "Stochastic decomposition in the M/G/1 queue with generalized vacations," *Operations Research*, vol. 33, pp. 1117–1129, 1985.

[33] J. Shanthikumar, "On stochastic decomposition in M/G/1 type queues with generalized server vacations," *Operations Research*, vol. 36, pp. 566–569, 1986.

[34] B. Doshi, "Queueing systems with vacations: A survey," *Queueing Systems*, vol. 1, no. 1, pp. 29–66, 1986.

[35] ——, "Single server queues with vacations," in *Stochastic Analysis of Computer and Communication Systems*, H. Takagi, Ed., 1990, pp. 217–265.

[36] J. R. Artalejo and G. I. Falin, "Stochastic decomposition for retrial queues," *TOP*, vol. 2, no. 2, pp. 329–342, 1994.

[37] B. K. Kumar and D. Arivudainambi, "The M/G/1 retrial queue with bernoulli schedules and general retrial times," *Computers and Mathematics with Applications*, vol. 43, pp. 15–30, 2002.

[38] P. Moreno, "An M/G/1 retrial queue with recurrent customers and general retrial times," *Applied Mathematics and Computation*, vol. 159, pp. 651–666, 2004.

[39] N. Sherman and J. Kharoufeh, "An M/M/1 retrial queue with unreliable server," *Operations Research Letters*, vol. 34, pp. 697–705, 2006.

[40] J. R. Artalejo, "New results in retrial queueing systems with breakdown of the servers," *Statistica Neerlandica*, vol. 48, no. 1, pp. 23–36, 2008.

[41] G. I. Falin, "The M/M/1 retrial queue with retrials due to server failures," *Queueing Systems*, vol. 58, pp. 155–160, 2008.

[42] J. Wang, "On the single server retrial queue with priority subscribers and server breakdowns," *Journal of System Science & Complexity*, vol. 21, pp. 304–315, 2008.

[43] N. Sherman, J. Kharoufeh, and M. Abramson, "An M/G/1 retrial queue with unreliable server for streaming multimedia applications," *Probability in the Engineering and Informational Sciences*, vol. 23, no. 02, pp. 281–304, 2009.

[44] G. I. Falin, "An M/G/1 retrial queue with an unreliable server and general repair times," *Performance Evaluation*, vol. 67, pp. 569–582, 2010.

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

[46] B. D'Auria, "M/M/∞ queue with on-off service speeds," in *Proceedings of XXV International Seminar on Stability Problems for Stochastic Models, Maiori (SA), Italy*, 2005, pp. 131–137.

[47] ——, “Stochastic decomposition of the M/G/∞ queue in a random environment,” *Operations Research Letters*, vol. 35, pp. 805–812, 2007.

[48] I. Eliazar, “On the discrete-time G/GI/∞ queue,” *Probability in the Engineering and Informational Sciences*, vol. 22, no. 04, pp. 557–585, 2008.

[49] B. Fralix and I. Adan, “An infinite-server queue influenced by a semi-Markovian environment,” *Queueing Systems*, vol. 61, no. 1, pp. 65–84, 2009.

[50] M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions*. National Bureau of Standards, Washington D.C., 1964.

[51] L. Slater, “Confluent hypergeometric functions,” in *Handbook of Mathematical Functions*, M. Abromowitz and I. Stegun, Eds., 1964, pp. 503–515.

[52] D. Fitzgerald, “Tricomi and Kummer functions in occurrence, waiting time and exceedance statistics,” *Stochastic Environmental Research and Risk Analysis*, vol. 16, pp. 207–224, 2002.

[53] J. Keilson, J. Cozzolino, and H. Young, “A service system with unfilled requests repeated,” *Operations Research*, vol. 16, no. 6, pp. 1126–1137, 1968.

[54] G. Jonin and J. Sedol, “Telephone systems with repeated calls,” in *Proceedings of the 6th International Teletraffic Congress (ITC70)*, 1970, pp. 435–1.

M. Baykal-Gürsoy is an Associate Professor of Industrial and Systems Engineering at Rutgers, the State University of New Jersey. She received her Ph.D. in Systems Engineering at the University of Pennsylvania in 1988. She served in the Editorial Boards of *the Institute of Computer Sciences, Social-Informatics and Telecommunications Engineering (ICST) Transactions on Network Optimization and Control*, and *the International Journal of Operations Research and Information Systems (IJORIS)*. Her research interests are in the areas of stochastic modelling, queueing, Markov decision processes, stochastic games, with applications to traffic, security, and supply chains.

Z. Duan earned his Ph.D. degree in Industrial and Systems Engineering at Rutgers, the State University of New Jersey in 2012. His research interests include queueing, traffic modeling, and security games.