Performance Evaluation 83–84 (2015) 1–15

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

A single server queue with Markov modulated service rates and impatient customers

Bara Kim [a], Jeongsim Kim [b,*]

[a] Department of Mathematics, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, 136-701, Republic of Korea

[b] Department of Mathematics Education, Chungbuk National University, 52 Naesudong-ro, Seowon-gu, Cheongju, Chungbuk, 361-763, Republic of Korea

ARTICLE INFO

Article history:
Received 4 December 2013
Received in revised form 20 July 2014
Accepted 10 November 2014
Available online 21 November 2014

Keywords:
Impatient customer
Markov modulated service rate
Loss probability
Waiting time distribution
System size distribution

ABSTRACT

We consider a single server queue in which the customers wait for service for a fixed time and leave the system if the service has not begun within that time. The customers arrive according to a Poisson process and each arriving customer brings in a certain amount of phase-type distributed work. The service rate of a server varies according to the underlying continuous time Markov process with finite states. We construct a Markov process by using the age process and then obtain the stationary distribution of the Markov process. By using the results of the stationary distribution of the Markov process, we obtain the loss probability, the waiting time distribution and the system size distribution.

1. Introduction

In many service systems, customers wait for service for a limited time only and leave the system if not served during that time. Such customers with limited waiting time are usually referred to as impatient customers. Impatience phenomena are often encountered in real-time communication systems, inventory systems with storage of perishable goods, telecommunication networks, call centers, etc. (See [1–3].)

We consider a single server queue with varying service rates and impatient customers. There are several studies on queues with varying service rates. Boxma and Kurkova [4] considered the $M/G/1$ queue where the speed of the server alternates between two values; high speed periods with an exponential distribution and low speed periods with a general distribution. Baykal-Gursoy and Xiao [5] considered the $M/M/\infty$ queue subject to random interruptions with the durations being exponentially distributed. Queija [6] considered the $M/M/1$ processor sharing queue where the server changes its service rate according to a finite birth and death process. Zhou and Gans [7] considered the $M/M/1$ queue where the server changes its service rates only when a customer completes service and the server takes only two values of the service rates. Mahabhashyam and Gautam [8] studied the $M/G/1$ queue where the server speed changes according to a continuous time Markov process. Huang and Lee [9] studied the $M/G/1$ queue where the server speed changes according to a two-state Markov chain.

In regards to queues with impatient customers, there has been a large amount of research. Barrer [10,11] analyzed the $M/M/1$ and $M/M/c$ queues with constant impatience times and obtained the stationary queue size distributions. Jurkevic [12] analyzed the $M/M/c$ queue where the impatience time is the minimum of a constant and an exponentially distributed time. Jurkevic [13] studied the $M/M/c$ queue with general impatience times. Baccelli and Hebuterne [1] obtained

* Corresponding author. Tel.: +82 43 261 2716; fax: +82 43 275 2715.
E-mail addresses: bara@korea.ac.kr (B. Kim), jeongsimkim@chungbuk.ac.kr (J. Kim).

http://dx.doi.org/10.1016/j.peva.2014.11.002
0166-5316/

the waiting time distributions in the $M/M/c$ and $M/G/1$ queues with general impatience times. Boxma and de Waal [14] developed several approximations for the loss probability in the $M/M/c$ queue with general impatience times. Brandt and Brandt [15,16] considered the $M(n)/M(n)/c$ queue with general impatience times where the arrival and service rates are allowed to depend on the number n of calls in the system. For the $M(n)/M/c$ queue with general impatience times, see Movaghar [17]. de Kok and Tijms [18] and Xiong et al. [19] studied the $M/G/1$ queue with constant impatience time. de Kok and Tijms [18] obtained an expression for the distribution function of the workload in terms of the workload in the modified system. Xiong et al. [19] set up an integral equation for the workload distribution using level crossing analysis. An analytical solution for this equation was given only for $M/H_2/1+D$ queue. de Kok and Tijms [18] and Xiong et al. [19] presented approximations for the loss probability and the mean waiting time in the $M/G/1$ queue with constant impatience time.

As yet, there are no known results on the queues with Markov modulated service rate and impatient customers. We consider a single server queue with Markov modulated service rate and constant impatience time. Customers arrive according to a Poisson process and each arriving customer brings in a certain amount of phase-type distributed work. The service rate of a server varies according to a random environment process that is governed by a continuous time Markov process with finite states. The service discipline is first come, first-served. Each arriving customer enters the system, but is only willing to wait in the queue for a fixed time $\tau > 0$. A customer who waits for time τ without his service having begun leaves the system after that time τ and becomes a lost customer. Our model focuses on the impatient customer model with limitation on the waiting time (impatience until the beginning of service) and unaware customers, which is one of Baccelli et al.'s classification [20,1] of queueing systems with impatient customers.

Previous papers on queues with impatient customers were analyzed mainly by using the unfinished work. The Markov fluid queues can be used for the analysis of the unfinished work. Refer to Refs. [21–24] regarding to Markov fluid queues. Yazici and Akar [25] also used the Markov fluid queues to analyze the impatient customer model with limitation on the sojourn time (waiting time plus service time).

Because in our queueing model the service rate changes according to an environment process, the method using the unfinished work is very complicated to apply to our queueing model. Therefore, we use the elapsed waiting time of the customer at the head of the queue (i.e., the age of the oldest waiting customer) to construct a Markov process. Choi et al. [26] and Houdt and Blondia [27] used this type of age process to construct a Markov process. The age process has been used in the analysis of the queueing systems, refer to He [28].

The main contribution of this paper is the analysis of the queueing system with Markov modulated service rate and impatient customers. We construct a simple Markov process by using the age process and then obtain the stationary distribution of the Markov process. By using the results of the stationary distribution of the Markov process, we also obtain performance measures such as the loss probability, the waiting time distribution and the system size distribution.

This paper is organized as follows. In Section 2, we describe the model in detail. In Section 3, we construct a simple Markov process and give a differential equation for the stationary distribution of the Markov process. Then we obtain the stationary distribution of the Markov process by solving the equation. In Section 4, we obtain the loss probability, the waiting time distribution and the system size distribution. In Section 5, numerical examples are presented to illustrate the result. Appendix is devoted to the derivation of useful properties which are used in Section 3.

2. Model description

We consider a single server queue with Markov modulated service rate and constant impatience time τ. Customers arrive according to a Poisson process with rate λ and each arriving customer brings in a certain amount of work. The amount of work has a phase-type distribution with representation $(\boldsymbol{\alpha}, T)$ of order m. The service rate of a server varies according to the environment process $\{J(t) : t \geq 0\}$, which is assumed to be irreducible Markov process with finite state space $\{1, \dots, n\}$ and infinitesimal generator $Q = (Q_{ij})_{1\leq i,j\leq n}$. When the state of the environment process $J(t) = i$, the service rate is μ_i, $i = 1, \dots, n$. That is, the server can serve μ_i amount of work per unit time.

Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$ be the stationary probability vector of the underlying Markov process $\{J(t) : t \geq 0\}$, i.e., $\boldsymbol{\pi}$ is a unique solution of $\boldsymbol{\pi}Q = \mathbf{0}$ and $\boldsymbol{\pi}\mathbf{1}_n = 1$, where $\mathbf{1}_n$ is the n-dimensional column vector with all components equal to one. Throughout the paper, $\mathbf{0}$ denotes the zero (row or column) vector of appropriate size determined by the context. The fundamental service rate is given by $\mu = \boldsymbol{\pi}M\mathbf{1}_n$, where $M = \mathrm{diag}(\mu_1, \dots, \mu_n)$, and the offered load ρ is defined as $\rho = \frac{\lambda\boldsymbol{\alpha}(-T)^{-1}\mathbf{1}_m}{\mu}$.

We close this section with a brief review of phase-type distribution. Consider a continuous-time Markov process with state space $\{1, \dots, m, m+1\}$ and an infinitesimal generator of the form

$$\begin{bmatrix} T & \boldsymbol{T}^0 \\ \mathbf{0} & 0 \end{bmatrix}, \tag{1}$$

where $T = (T_{ij})_{1\leq i,j\leq m}$ is a nonsingular $m \times m$ matrix and $\boldsymbol{T}^0 = (T_1^0, \dots, T_m^0)^\top$ is an m-dimensional column vector satisfying $T\mathbf{1}_m + \boldsymbol{T}^0 = \mathbf{0}$. Each of the states of the Markov process represents one of the phases. Let $(\boldsymbol{\alpha}, 0)$ be the initial distribution of the Markov process, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$ with $\sum_{i=1}^m \alpha_i = 1$. Then the time until absorption into the state $m+1$ in the Markov process has the phase-type distribution with representation $(\boldsymbol{\alpha}, T)$ and mean $-\boldsymbol{\alpha}T^{-1}\mathbf{1}_m$. Without loss of generality,