Journal of Computational and Applied Mathematics 236 (2012) 3445–3460

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

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

Tail asymptotics of the queue size distribution in the M/M/m retrial queue[☆]

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

[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, Heungdeok-gu, Cheongju, Chungbuk, 361-763, Republic of Korea

ARTICLE INFO

Article history:
Received 20 April 2010
Received in revised form 31 January 2012

Keywords:
M/M/m retrial queue
Queue size distribution
Censored Markov process
Tail asymptotics
Karamata Tauberian theorem
Riemann–Lebesgue lemma

ABSTRACT

We consider an M/M/m retrial queue and investigate the tail asymptotics for the joint distribution of the queue size and the number of busy servers in the steady state. The stationary queue size distribution with the number of busy servers being fixed is asymptotically given by a geometric function multiplied by a power function. The decay rate of the geometric function is the offered load and independent of the number of busy servers, whereas the exponent of the power function depends on the number of busy servers. Numerical examples are presented to illustrate the result.

1. Introduction

Retrial queues are queueing systems in which arriving customers who find all servers occupied may retry for service again after a random amount of time. Retrial queues have been widely used to model many problems in telephone systems, call centers, telecommunication networks, computer networks and computer systems, and in daily life. Detailed overviews for retrial queues can be found in the bibliographies [1–3], the surveys [4–6], and the books [7,8].

In this paper we consider an M/M/m retrial queue where customers arrive from outside the system according to a Poisson process with rate λ. The service facility consists of m identical servers, and service times are exponentially distributed with mean μ^{-1}. If there is a free server when a customer arrives from outside the system, this customer begins to be served immediately and leaves the system after the service is completed. On the other hand, any customer who finds all the servers busy upon arrival joins a retrial group, called an orbit, and then attempts service after a random amount of time. If there is a free server when a customer from the orbit attempts service, this customer receives service immediately and leaves the system after the service completion. Otherwise the customer comes back to the orbit immediately and repeats the retrial process. The retrial time, i.e., the length of the time interval between two consecutive attempts made by a customer in the orbit, is exponentially distributed with mean ν^{-1}. The arrival process, the service times, and the retrial times are assumed to be mutually independent. The traffic load ρ is defined as $\rho = \frac{\lambda}{m\mu}$. It is known that the M/M/m retrial queue is stable if and only if $\rho < 1$ [9]. We assume that $\rho < 1$ for stability of the system.

The M/M/m retrial queue has been studied by several authors. Greenberg and Wolff [10] studied an approximation method for steady-state probability and provided an upper bound on system performance for the M/M/m retrial queue

☆ The second author's research was supported by the Korea Research Foundation (KRF) grant funded by the Korea government (MEST) (2009-0076674). The third author's research was supported by the Korea Research Foundation (KRF) grant funded by the Korea government (MEST) (2009-0076600).

* Corresponding author.
E-mail addresses: b1155@hanmail.net (J. Kim), jeongsimkim@chungbuk.ac.kr (J. Kim), bara@korea.ac.kr (B. Kim).

0377-0427/$ – see front matter
doi:10.1016/j.cam.2012.03.027

with a finite capacity. Pearce [11] found the joint distribution of the number of customers in the orbit and the number of busy servers for the M/M/m retrial queue with the orbit of finite capacity. For the M/M/m retrial queue, Neuts and Rao [12] presented a simplifying approximation of the joint distribution of the number of customers in the orbit and the number of busy servers at steady state by placing a fictitious limit in the orbit capacity. This approximation based on truncation method, starting from the pioneering paper in [13], has been widely used in the numerical analysis of retrial queues.

Analytical studies of the multiserver retrial models are limited and in many cases restricted to two-server systems. Keilson et al. [14] established a recursive algorithm for the computation of steady-state probabilities in the M/M/2 retrial queue. Hanschke [15] showed that the generating functions of the steady-state probabilities can be expressed in terms of generalized hypergeometric functions in the M/M/2 retrial queue.

The main contribution of our work is that we find the tail asymptotics for the distribution of the queue size (i.e., the number of customers in the orbit) in the M/M/m retrial queue. Tail behaviors of the queue size and the waiting time distributions in retrial queues began to be investigated recently. Nobel and Tijms [16] and Kim et al. [17] studied light-tailed asymptotic behaviors in the M/G/1 retrial queue where the service time distribution has a finite exponential moment. Nobel and Tijms [16] suggested a light-tailed approximation of the waiting time distribution. Kim et al. [17] showed that the tail of the queue size distribution is asymptotically given by a geometric function multiplied by a power function. Kim et al. [18] extended the result of [17] to the case of MAP/G/1 retial queue. On the other hand, Shang et al. [19] and Kim and Kim [20] studied heavy-tailed asymptotics in the M/G/1 retrial queue. Shang et al. [19] showed that the stationary distribution of the queue size in the M/G/1 retrial queue is subexponential if the stationary distribution of the queue size in the corresponding ordinary M/G/1 queue is subexponential. As a corollary of this property, they proved that the stationary distribution of the queue size has a regularly varying tail if the service time distribution has a regularly varying tail. Kim and Kim [20] showed that if the service time distribution has a regularly varying tail of index $-\alpha$, $\alpha > 1$, then the waiting time distribution has a regularly varying tail of index $1 - \alpha$.

In this paper we investigate the tail asymptotics for the joint distribution of the number of customers in the orbit and the number of busy servers at steady state in the M/M/m retrial queue. More precisely, we show that for $i = 0, 1, \dots, m$,

$$\mathbb{P}(N = n, S = i) \sim \frac{c}{i!}\left(\frac{\nu}{\mu}\right)^i n^{\frac{\lambda}{m\nu} - m + i} \rho^n \quad \text{as } n \to \infty,$$

where N is the number of customers in the orbit at steady state, S is the number of busy servers at steady state, and c is the positive constant. As shown in the above formula, the stationary queue size distribution with the number of busy servers being fixed, is asymptotically given by a geometric function multiplied by a power function. The decay rate of the geometric function is the offered load ρ and independent of the number of busy servers, whereas the exponent of the power function depends on the number of busy servers.

In order to derive the main result, we first consider a censored Markov process obtained by observing the M/M/m retrial queue only when the number of busy servers is less than or equal to $m - 1$. A matrix differential equation is derived for the vector probability generating function of the stationary distribution of the censored Markov process. The result is obtained by studying analytic properties of the solution of the differential equation.

The remainder of this paper is organized as follows. In the next section, we briefly review our model and introduce notations. In Section 3, we present our main result on the tail asymptotics of the queue size distribution. Section 4 is devoted to the derivation of the tail asymptotics stated without proof in Section 3. In Section 5, numerical examples are presented to illustrate the result. In Appendix, we prove the results stated without proof in Section 4.

2. The M/M/m retrial queue

We consider the M/M/m retrial queue as described in Section 1. By a little abuse of notation, let $N(t)$ denote the number of customers in the orbit at time t and $S(t)$ the number of busy servers at time t. Then $\{(N(t), S(t)) : t \geq 0\}$ is a continuous time Markov process with state space $\{0, 1, 2, \dots\} \times \{0, 1, \dots, m\}$. The infinitesimal generator Q of $\{(N(t), S(t)) : t \geq 0\}$ is given by

$$Q = \begin{bmatrix} B_0 & A_0 & O & O & \cdots \\ C_1 & B_1 & A_1 & O & \cdots \\ O & C_2 & B_2 & A_2 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix},$$

where $A_n, B_n,\ n \geq 0$ and $C_n,\ n \geq 1$, are $(m+1) \times (m+1)$ matrices whose (i, j) components $(A_n)_{ij}$, $(B_n)_{ij}$ and $(C_n)_{ij}$, $0 \leq i,\ j \leq m$ are given by

$$(A_n)_{ij} = \begin{cases} \lambda & \text{if } i = j = m, \\ 0 & \text{otherwise,} \end{cases}$$