A discrete-time Geo/G/1 retrial queue with starting failures and second optional service

Jinting Wang*, Qing Zhao

Department of Mathematics, Beijing Jiaotong University, Beijing 100044, PR China

Received 1 December 2005; received in revised form 10 October 2006; accepted 11 October 2006

Abstract

We consider a discrete-time Geo/G/1 retrial queue with starting failures in which all the arriving customers require a first essential service while only some of them ask for a second optional service. We study the Markov chain underlying the considered queuing system and its ergodicity condition. Explicit formulae for the stationary distribution and some performance measures of the system in steady state are obtained. We also obtain two stochastic decomposition laws regarding the probability generating function of the system size. Finally, some numerical examples are presented to illustrate the influence of the parameters on several performance characteristics.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Discrete-time retrial queues; Essential and optional service; Repair; Stochastic decomposition; Unreliable server

1. Introduction

There is a growing interest in the analysis of discrete-time queues due to their applications in communication systems and other related areas. Many computer and communication systems operate on a discrete-time basis where events can only happen at regularly spaced epochs. One of the main reasons for analyzing discrete-time queues is that these systems are more appropriate than their continuous-time counterparts for modelling computer and telecommunication systems. In the last few years, a number of queueing models have been analyzed in discrete-time, details of which may be found in recent books [1–5].

Queueing systems with repeated attempts have been widely used to model many practical problems in telephone switching systems, telecommunication networks, computer and communication systems. For a detailed review of the main results and the literature on this topic the reader is referred to [6–8]. In the past, the study of the retrial queues has been focused on the continuous case, but recently Yang and Li [9] have extended the study to discrete-time retrial queues. In spite of the importance of the discrete-time retrial queues, little is still known about them. In fact, only very

* Research sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry and the National Natural Science Foundation of China (Nos. 10526004, 60504016).

* Corresponding author. Fax: +86 10 51688433.

E-mail addresses: jtwang@bjtu.edu.cn, wjting@yahoo.com (J. Wang).
few people have analyzed discrete-time retrial queues [9–16]. This fact gives the reason why efforts should be taken to fill this gap.

Recently, there have been several contributions considering queueing systems of $M/G/1$ type in which the server may provide a second phase of service. Such queueing situations occur in day-to-day life, where all the arriving customers require the main service and only some may require the subsidiary service provided by the same server. Madan [17] studies an $M/G/1$ queue with second optional service in which the first essential service time follows a general distribution but the second optional service is assumed to be exponentially distributed. Some examples of queueing situations where such service mechanism can arise are also given. Medhi [18] generalizes the model by considering that the second optional service is also governed by a general distribution. Wang [19] considers the model with the assumption that the server is subject to breakdowns and repairs, and some crucial reliability and queueing indexes are obtained. Atencia and Moreno [20] extend the study to the discrete-time retrial queues with second optional service but without server breakdowns.

It should be noted that in most retrial queueing literature the server is assumed to be available on a permanent basis. Nevertheless, these assumptions are practically unrealistic. Due to the limited ability of repairs and the heavy influence of the breakdowns on the performance measure of the system, it is important to study retrial queues with server breakdowns and repairs in many practical situations. Most of the existing works focus on continuous-time models, see [21–27] and the references therein. In [27], the authors summarize that there are four classes of important results regarding the queues with server subject to breakdowns and repairs during the service. However, in this paper we assume the server is subject to starting failures, which is different from the four classes indicated in [27]. That is, we assume that an arriving customer (new or returning) must turn on the server when the server is idle upon arrival, which takes zero time. If the server is started successfully, the customer gets service immediately. Otherwise, the server undergoes “repair” immediately and the customer must leave and make a retrial at a later time. The server is assumed to be reliable during service. Yang and Li [24] studied the $M/G/1$ retrial queue with starting failures in continuous time for the first time, and later Krishna Kumar et al. [28] studied the $M/G/1$ retrial queue with starting failures and feedback, where the retrial time is assumed to follow an arbitrary distribution and the customers in the orbit access the server under FCFS discipline. However, there is very little work that deal with discrete-time retrial queues incorporating unreliable server (see [29]). This motivates us to investigate such queueing systems in this work.

In the paper, we consider a discrete-time Geo/$G/1$ retrial queue with starting failures in which some customers may demand a second service in addition to the first essential service. The contribution of this work is that we develop the methodology used in [24] to deal with discrete-time retrial queue with unreliable server and optional service. We extend the analysis of unreliable retrial queues in continuous-time to the discrete-time counterpart. It can be shown that the models studied in [20,24] are special cases of our model.

The remainder of the paper is organized as follows. In the next section, the mathematical description of the model is introduced. In Section 3, we analyze the Markov chain associated with the system and the stability condition of the system. The orbit and system sizes distribution are obtained together with several performance measures of the system. Section 4 gives two different stochastic decomposition laws regarding the probability generating function of the system size. Finally, in Section 5, we present some numerical results to illustrate the impact of the service rate, retrial rate and unreliability factor on the performance of the system.

2. Model description

Let us consider a discrete-time single server retrial queue where the time axis is segmented into a sequence of equal time intervals (called slots). Further, let the time axis be marked by $0, 1, \ldots, m, \ldots$. Different from continuous-time queues, the probability of an arrival and a departure and other queueing activities occurring concurrently may be not zero in discrete-time queues anymore. So it is necessary to specify the order in which the arrivals and departures take place in case of simultaneity. In literature, there are two policies: (i) if an arrival takes precedence over a departure, then it is known as a late arrival system (LAS), and (ii) if a departure takes precedence over an arrival, then it is referred to as an early arrival system (EAS). They are also known as arrival first (AF) and departure first (DF) policies, respectively (see [2]). In the present paper, we discuss the case (ii), that is, we discuss the model only for early arrival system (EAS). For mathematical clarity, we assume that the departures and the end of the repairs occur in the interval $(m^{-}, m)$, and the arrivals, the retrials and the beginning of the repairs occur in the interval $(m, m^{+})$; that is, the arrivals, the retrials and the beginning of the repairs occur at the moment immediately after the slot boundaries, and
the departures and the end of the repairs occurs at the moment immediately before the slot boundaries. One can study the same model with LAS assumption and different results will be obtained.

Customers arrive according to a Bernoulli arrival process with probability \( p \). If, upon arrival, the server is busy or down, the customer is obliged to leave the service area and to come back to the system after a random amount of time (in slots). Those customers who are waiting to retry services are considered to be in orbit. An arriving (external or repeated) customer who finds the server idle must turn on the service station. If the server is activated successfully (with a probability \( \theta \)), the customer begins his service immediately; otherwise, if the server is started unsuccessfully (with a complementary probability \( \bar{\theta} \)), the server is sent to repair directly and the customer must join the orbit. Obviously, this failure/repair assumption is different from that in [26, 27], which has been explained in the ‘Introduction’ section. On completion of the first essential service, a customer decides with probability \( \alpha \) to receive a second optional service and with complementary probability \( \bar{\alpha} \) to abandon the system forever.

Each customer in the orbit forms an independent retrial source and the retrial time (the time between two successive attempts by the same customer) follows a geometrical law with probability \( 1 - r \). The retrial process of a repeated customer concludes only if, upon one particular attempt, the server is idle and the repeated customer is chosen for the service among all other repeated customers who are attempting the service at that time and the service station is activated successfully.

The service times of the 1st (essential) and 2nd (optional) services are independent and arbitrarily distributed with distributions \( \{s_{1,i}\}_{i=1}^{\infty} \) and \( \{s_{2,j}\}_{j=1}^{\infty} \), and probability generating functions \( S_1(x) = \sum_{i=1}^{\infty} s_{1,i}x^i \) and \( S_2(x) = \sum_{j=1}^{\infty} s_{2,j}x^j \), respectively. The corresponding \( n \)th factorial moments will be denoted by \( \beta_{1,n} \) and \( \beta_{2,n} \).

The repair times are independent and identically distributed with arbitrary distribution \( \{s_{3,i}\}_{i=1}^{\infty} \), generating function \( S_3(x) = \sum_{i=1}^{\infty} s_{3,i}x^i \) and \( n \)th factorial moments \( \beta_{3,n} \). We assume that after repair the server is as good as new.

Finally, it is assumed that the inter-arrival times, the retrial times, the service times, and the repair times are mutually independent. Further, in order to avoid trivial cases, we suppose \( 0 < p < 1, 0 \leq r < 1 \) and \( 0 < \theta \leq 1 \). We also denote by \( \bar{p} \equiv 1 - p \) and \( \rho \equiv \rho_1 + \rho_2 + \rho_3 \) the traffic intensity, where \( \rho_1 = p\rho_{1,1}, \rho_2 = \alpha p\rho_{2,1} \) and \( \rho_3 = \frac{\bar{\theta}}{p} \rho_{3,1} \) denote the system load due to arrivals in the first essential service, the second optional service, and the repair process caused by starting failures, respectively. To explain \( \rho_3 \), we consider an arbitrary customer and assume that, at a particular visit, it finds the server idle and causes the server to undergo repair. Since its success probability is always \( \theta \), the total number of such visits to the server before receiving service has a geometric distribution with mean \( \frac{\bar{\theta}}{p} \). Recall that \( \beta_{3,1} \) is the mean repairing time. Thus, \( \frac{\bar{\theta}}{p}\beta_{3,1} \) is the expected total repairing time triggered by this particular customer, and \( \rho_3 \) is the system load due to this factor.

3. The Markov chain

At time \( m^+ \), the system can be described by the process
\[
X_m = (C_m, \xi_m, N_m)
\]
where \( C_m \) denotes the state of the server (0, 1, 2 or 3 according to whether the server is free, busy providing a first essential, busy providing a second optional service or down) and \( N_m \), the number of customers in the retrial group. If \( C_m \in \{1, 2\} \), then \( \xi_m \) represents the remaining service time of the customer currently being served and if \( C_m = 3 \), \( \xi_m \) corresponds to the remaining repair time.

After introducing the above supplementary variables, the future dynamics of \( X_m \) depends only on the current state. Or, given the current state, the next state and the evolution of the system prior to the current state are independent. So it can be shown that \( \{X_m, m \in \mathbb{N}\} \) is the Markov chain of our queueing system, whose state space is
\[
S = \{(0, k) : k \geq 0; (j, i, k) : j = 1, 2, i \geq 1, k \geq 0; (3, i, k) : i \geq 1, k \geq 1\}.
\]

Our objective is to find the stationary distribution
\[
\pi_{0,k} = \lim_{m \to \infty} P[C_m = 0, N_m = k]; \quad k \geq 0,
\]
\[
\pi_{j,i,k} = \lim_{m \to \infty} P[C_m = 1, \xi_m = i, N_m = k]; \quad i \geq 1, k \geq 0, \quad j = 1, 2,
\]
\[
\pi_{3,i,k} = \lim_{m \to \infty} P[C_m = 2, \xi_m = i, N_m = k]; \quad i \geq 1, \quad k \geq 1,
\]
of the Markov chain \( \{X_m, m \in \mathbb{N}\} \).

The one-step transition probabilities are given by:

1. if \( k \geq 0 \)
   \[
   \begin{align*}
   p(0,k)(0,k) &= \tilde{\rho} r^k, \\
   p(1,1,k)(0,k) &= \tilde{\alpha} \tilde{\rho} r^k, \\
   p(2,1,k)(0,k) &= \tilde{\rho} r^k, \\
   p(3,1,k)(0,k) &= \tilde{\rho} r^k, \quad (k \geq 1)
   \end{align*}
   \]

2. if \( i \geq 1, k \geq 0 \)
   \[
   \begin{align*}
   p(0,k)(i,k) &= p \theta s_{1,i}, \\
   p(0,k+1)(i,k) &= \tilde{p}(1 - r^{k+1}) \theta s_{1,i}, \\
   p(1,1,k)(i,k) &= \tilde{\alpha} p \theta s_{1,i}, \\
   p(1,1,k+1)(i,k) &= \tilde{\alpha} \tilde{p}(1 - r^{k+1}) \theta s_{1,i}, \\
   p(1,i+1,k)(i,k) &= p, \quad (k \geq 1) \\
   p(2,i+1,k)(i,k) &= \tilde{p}, \\
   p(2,1,k)(i,k) &= p \theta s_{1,i}, \\
   p(2,1,k+1)(i,k) &= \tilde{p}(1 - r^{k+1}) \theta s_{1,i}, \\
   p(3,1,k)(i,k) &= p \theta s_{1,i}, \quad (k \geq 1) \\
   p(3,1,k+1)(i,k) &= \tilde{p}(1 - r^{k+1}) \theta s_{1,i},
   \end{align*}
   \]

3. if \( i \geq 1, k \geq 0 \)
   \[
   \begin{align*}
   p(1,k-1)(2,i,k) &= \alpha s_{2,i} p, \quad (k \geq 1) \\
   p(1,1,k)(2,i,k) &= \alpha s_{2,i} \tilde{p}, \\
   p(2,i+1,k-1)(2,i,k) &= p, \quad (k \geq 1) \\
   p(2,i+1,k)(2,i,k) &= \tilde{p},
   \end{align*}
   \]

4. if \( i \geq 1, k \geq 1 \)
   \[
   \begin{align*}
   p(0,k-1)(3,i,k) &= p \tilde{\theta} s_{3,i}, \\
   p(0,k)(3,i,k) &= \tilde{p}(1 - r^k) \tilde{\theta} s_{3,i}, \\
   p(1,k-1)(3,i,k) &= \tilde{\alpha} p \tilde{s}_{3,i}, \\
   p(1,1,k)(3,i,k) &= \tilde{\alpha} \tilde{p}(1 - r^k) \tilde{s}_{3,i}, \\
   p(2,1,k-1)(3,i,k) &= p \tilde{\theta} s_{3,i}, \\
   p(2,1,k)(3,i,k) &= \tilde{p}(1 - r^k) \tilde{\theta} s_{3,i}, \\
   p(3,1,k-1)(3,i,k) &= p \tilde{\theta} s_{3,i}, \quad (k \geq 2) \\
   p(3,1,k)(3,i,k) &= \tilde{p}(1 - r^k) \tilde{\theta} s_{3,i}, \\
   p(3,i+1,k-1)(3,i,k) &= p, \quad (k \geq 2) \\
   p(3,i+1,k)(3,i,k) &= \tilde{p},
   \end{align*}
   \]

where \( \tilde{\rho} = 1 - p \).

According to the above transition probabilities, it is easy to write down the transition probability matrix which will be of the M/G/1 type. However, we do not need to analyze the matrix in the sequel, hence it is omitted here.

The Kolmogorov equations for the stationary distribution are
\[
\pi_{0,k} = \tilde{\rho} r^k \pi_{0,k} + \tilde{\alpha} \tilde{\rho} r^k \pi_{1,1,k} + \tilde{\rho} r^k \pi_{2,1,k} + (1 - \delta_{0k}) \tilde{p} r^k \pi_{3,1,k}, \quad k \geq 0, \tag{3.1}
\]
\[ \pi_{1,i,k} = p \theta s_{1,i,1} \pi_{0,k} + \bar{p}(1 - r^{k+1}) \theta s_{1,i,1} \pi_{1,k+1} + \bar{a} p \theta s_{1,i,1} \pi_{1,k} + \bar{a} \bar{p}(1 - r^{k+1}) s_{1,i,1} \pi_{1,k+1} + (1 - \delta_0) p \theta s_{1,i,1} \pi_{1,k+1} + \bar{p}(1 - r^{k+1}) s_{1,i,1} \pi_{2,k+1} + \bar{p}(1 - r^{k+1}) \theta s_{1,i,1} \pi_{2,k+1}, \quad i \geq 1, k \geq 0, \]  
\[ \pi_{2,i,k} = (1 - \delta_0) \alpha s_{2,i} p \pi_{1,k+1} + \alpha s_{2,i} \bar{p} \pi_{1,k+1} + (1 - \delta_0) p \pi_{2,i,k+1} + \bar{p} \pi_{2,i,k+1}, \quad i \geq 1, k \geq 0, \]  
\[ \pi_{3,i,k} = p \bar{\theta} s_{3,i,1} \pi_{0,k-1} + \bar{p}(1 - r^{k}) \bar{\theta} s_{3,i,1} \pi_{k} + \bar{a} p \bar{\theta} s_{3,i,1} \pi_{k-1} + \bar{a} \bar{p}(1 - r^{k}) \bar{\theta} s_{3,i,1} \pi_{k-1} + \bar{\theta} s_{3,i,1} \pi_{2,k-1} + \bar{p}(1 - r^{k}) \bar{\theta} s_{3,i,1} \pi_{2,k-1} + (1 - \delta_1) p \bar{\theta} s_{3,i,1} \pi_{2,k-1} + \bar{p}(1 - r^{k}) \bar{\theta} s_{3,i,1} \pi_{3,k-1} + (1 - \delta_1) p \bar{\theta} s_{3,i,1} \pi_{3,k-1} + \bar{p} \pi_{3,i,k+1}, \quad i \geq 1, k \geq 1, \]  
and the normalization condition is
\[ \sum_{k=0}^{\infty} \pi_{0,k} + \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} \pi_{j,i,k} + \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \pi_{3,i,k} = 1. \]  

To solve this system of equations, we introduce the generating functions
\[ \phi_0(z) = \sum_{k=0}^{\infty} \pi_{0,k} z^k; \]
\[ \phi_j(x, z) = \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} \pi_{j,i,k} x^i z^k, \quad j = 1, 2; \]
\[ \phi_3(x, z) = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \pi_{3,i,k} x^i z^k; \]
and the auxiliary generating functions
\[ \phi_{j,i}(z) = \sum_{k=0}^{\infty} \pi_{j,i,k} z^k, \quad i \geq 1, \quad j = 1, 2; \]
\[ \phi_{3,i}(z) = \sum_{k=1}^{\infty} \pi_{3,i,k} z^k, \quad i \geq 1. \]

Multiplying (3.1)–(3.4) by \( z^k \) and summing over \( k \), these equations become
\[ \phi_0(z) = \bar{p} [\phi_0(rz) + \bar{a} \phi_{1,1}(rz) + \phi_{2,1}(rz) + \phi_{3,1}(rz)], \]
\[ \phi_{1,i}(z) = (\bar{p} + pz) \phi_{1,i+1}(z) + \bar{p} + pz \theta s_{1,i} [\phi_0(z) + \bar{a} \phi_{1,1}(z) + \phi_{2,1}(z) + \phi_{3,1}(z)] \]
\[ - \frac{\bar{p}}{z} \theta s_{1,i} [\phi_0(z) + \bar{a} \phi_{1,1}(z) + \phi_{2,1}(z) + \phi_{3,1}(z)], \]
\[ \phi_{2,i}(z) = \alpha s_{2,i} (\bar{p} + pz) \phi_{1,i}(z) + (\bar{p} + pz) \phi_{2,i+1}(z), \]
\[ \phi_{3,i}(z) = (\bar{p} + pz) \phi_{3,i+1}(z) + (\bar{p} + pz) \bar{\theta} s_{3,i} [\phi_0(z) + \bar{a} \phi_{1,1}(z) + \phi_{2,1}(z) + \phi_{3,1}(z)] \]
\[ - \frac{\bar{p}}{z} \bar{\theta} s_{3,i} [\phi_0(z) + \bar{a} \phi_{1,1}(z) + \phi_{2,1}(z) + \phi_{3,1}(z)]. \]

By substituting Eq. (3.5) into Eq. (3.6), (3.8), we get
\[ \phi_{1,i}(z) = (\bar{p} + pz) \phi_{1,i+1}(z) - \frac{1 - z}{z} p \theta s_{1,i} \phi_0(z) \]
\[ + \frac{\bar{p} + pz}{z} \theta s_{1,i} [\bar{a} \phi_{1,1}(z) + \phi_{2,1}(z) + \phi_{3,1}(z)], \]
\[ \phi_{3,i}(z) = (\bar{p} + pz) \phi_{3,i+1}(z) - (1 - z) p \bar{\theta} s_{3,i} \phi_0(z) \]
\[ + (\bar{p} + pz) \bar{\theta} s_{3,i} [\bar{a} \phi_{1,1}(z) + \phi_{2,1}(z) + \phi_{3,1}(z)]. \]
Then, multiplying (3.7), (3.9) and (3.10) by $x^i$ and summing over $i$ leads to
\[
\frac{x - (\bar{p} + pz)}{x} \phi_1(x, z) = \frac{\bar{p} + pz}{z} \theta S_1(x)[\phi_2,1(z) + \phi_3,1(z)] + \frac{\bar{p} + pz}{z} [\bar{\alpha} \theta S_1(x) - z] \phi_1,1(z) \\
- \left(1 - \frac{z}{z} \right) \frac{p \theta S_1(x) \phi_0(z)}{z}, \tag{3.11}
\]
\[
\frac{x - (\bar{p} + pz)}{x} \phi_2(x, z) = \alpha(\bar{p} + pz) S_2(x) \phi_1,1(z) - (\bar{p} + pz) \phi_2,1(z), \tag{3.12}
\]
\[
\frac{x - (\bar{p} + pz)}{x} \phi_3(x, z) = (\bar{p} + pz)[\bar{\theta} S_3(x) - 1] \phi_3,1(z) + (\bar{p} + pz) \bar{\theta} S_3(x)[\bar{\alpha} \phi_1,1(z) + \phi_2,1(z)] \\
- (1 - \frac{z}{z}) p \bar{\theta} S_3(x) \phi_0(z). \tag{3.13}
\]
Choosing $x = \bar{p} + pz$ in (3.11)–(3.13) yields
\[
(1 - z) p \theta S_1(\bar{p} + pz) \phi_0(z) = (\bar{p} + pz)[\bar{\alpha} \theta S_1(\bar{p} + pz) - z] \phi_1,1(z) \\
+ (\bar{p} + pz) \theta S_1(\bar{p} + pz)[\phi_2,1(z) + \phi_3,1(z)], \tag{3.14}
\]
\[
\alpha(\bar{p} + pz) S_2(\bar{p} + pz) \phi_1,1(z) = (\bar{p} + pz) \phi_2,1(z), \tag{3.15}
\]
\[
(1 - z) p \bar{\theta} S_3(\bar{p} + pz) \phi_0(z) = (\bar{p} + pz) \bar{\theta} S_3(\bar{p} + pz)[\bar{\alpha} \phi_1,1(z) + \phi_2,1(z)] \\
+ (\bar{p} + pz) [\bar{\theta} S_3(\bar{p} + pz) - 1] \phi_3,1(z). \tag{3.16}
\]
The following two lemmas, whose proofs can be readily obtained, will be used later on.

**Lemma 3.1.** The inequalities $S_1(x) \leq x$, $S_2(x) \leq x$ and $S_3(x) \leq x$ hold for $0 \leq x \leq 1$.

**Proof.** Because of the convex properties of the functions $S_1(x)$, $S_2(x)$ and $S_3(x)$, the proof of this lemma is immediately obtained. \[ \Box \]

**Lemma 3.2.** If $\rho \equiv \rho_1 + \rho_2 + \rho_3 < 1$, then inequality $\theta S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)] + z \bar{\theta} S_3(\bar{p} + pz) - z > 0$ holds for $0 \leq z < 1$. Under this condition, the following limit exists:
\[
\lim_{z \to 1} \frac{\theta S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)] + z \bar{\theta} S_3(\bar{p} + pz) - z}{1 - z} = \frac{1}{\theta(1 - \rho_1 - \rho_2 - \rho_3)}.
\]

**Proof.** Consider the function $v(z) = \theta S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)] + z \bar{\theta} S_3(\bar{p} + pz)$. It is easy to see that if $\rho \equiv \rho_1 + \rho_2 + \rho_3 < 1$:

1. $v(0) = \theta S_1(\bar{p})[\bar{\alpha} + \alpha S_2(\bar{p})] > 0$;
2. $v(1) = 1$;
3. $v'(z) = \theta p \beta_{1,1} + \theta p \alpha \beta_{2,1} + \theta + \bar{p} \beta_{3,1} < 1$;
4. $v''(z) > 0$.

Noting that the existence of a $z_0 \in (0, 1)$ such that $v(z_0) = z_0$ leads to a contradiction with the above properties, thus the first part of Lemma 3.2 is proved. The rest of the lemma can be calculated based on the above inequality and the l’Hôpital’s rule. \[ \Box \]

From (3.14)–(3.16), we find the generating functions $\phi_{1,1}(z)$, $\phi_{2,1}(z)$ and $\phi_{3,1}(z)$ as follows:
\[
\phi_{1,1}(z) = \frac{S_1(\bar{p} + pz)}{\bar{p} + pz} \frac{p \theta (1 - z) \phi_0(z)}{\theta S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)] + z \bar{\theta} S_3(\bar{p} + pz) - z}, \tag{3.17}
\]
\[
\phi_{2,1}(z) = \frac{S_2(\bar{p} + pz)}{\bar{p} + pz} \frac{\alpha p \theta (1 - z) S_1(\bar{p} + pz) \phi_0(z)}{\theta S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)] + z \bar{\theta} S_3(\bar{p} + pz) - z}, \tag{3.18}
\]
\[
\phi_{3,1}(z) = \frac{S_3(\bar{p} + pz)}{\bar{p} + pz} \frac{p \bar{\theta} z (1 - z) \phi_0(z)}{\theta S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)] + z \bar{\theta} S_3(\bar{p} + pz) - z}. \tag{3.19}
\]
It should be noted that, by using Lemma 3.2, the above functions are defined for $0 \leq z < 1$ and in $z = 1$ can be extended by continuity if $\rho_1 + \rho_2 + \rho_3 < 1$. 

Inserting $\phi_{1,1}(r z), \phi_{2,1}(r z)$ and $\phi_{3,1}(z)$ into Eq. (3.5) leads to

\[
\phi_0(z) = \frac{\theta S_1(\tilde{\rho} + prz)[\tilde{\alpha} + \alpha S_2(\tilde{\rho} + prz)] + rz\tilde{\theta} S_3(\tilde{\rho} + prz) - rz(\tilde{\rho} + prz)}{\tilde{\rho} + prz}[\theta S_1(\tilde{\rho} + prz)[\tilde{\alpha} + \alpha S_2(\tilde{\rho} + prz)] + z\tilde{\theta} S_3(\tilde{\rho} + prz) - rz] \tilde{\phi}_0(rz) \\
= G(rz)\phi_0(rz).
\]

(3.20)

It is easy to check that, by using (3.20) recursively, $\phi_0(z) = \phi_0(0) \prod_{k=1}^{\infty} G(r^k z)$. The convergence of this infinite product is established in the following lemma.

**Lemma 3.3.** If $\rho_1 + \rho_2 + \rho_3 < 1$, the infinite product $\prod_{k=1}^{\infty} G(r^k z)$ converges.

**Proof.** We express $G(z)$ as

\[
G(z) = 1 + F(z)
\]

where

\[
F(z) = \frac{\tilde{\rho} + pz - \theta S_1(\tilde{\rho} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{\rho} + pz)] - z\tilde{\theta} S_3(\tilde{\rho} + pz)}{(\tilde{\rho} + pz)[\theta S_1(\tilde{\rho} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{\rho} + pz)] + z\tilde{\theta} S_3(\tilde{\rho} + pz) - z]} pz.
\]

Applying Lemma 3.2 and the clear inequalities

\[
\tilde{\rho} + pz - \theta S_1(\tilde{\rho} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{\rho} + pz)] - z\tilde{\theta} S_3(\tilde{\rho} + pz) \\
\geq \theta[\tilde{\rho} + pz - \theta S_1(\tilde{\rho} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{\rho} + pz)]] + \tilde{\theta}[\tilde{\rho} + pz - S_2(\tilde{\rho} + pz)] \geq 0
\]

valid for $0 \leq z \leq 1$, it can be easily shown that

\[
F(z) \geq 0 \quad \text{for } 0 \leq z \leq 1 \text{ if } \rho_1 + \rho_2 + \rho_3 < 1.
\]

Considering Eq. (3.21), the infinite product can be rewritten as

\[
\prod_{k=1}^{\infty} G(r^k z) = \prod_{k=1}^{\infty} [1 + F(r^k z)].
\]

(3.22)

It is well-known that infinite product in (3.22) converges if and only if the series $\sum_{k=1}^{\infty} F(r^k z)$ is convergent, which is obvious since $\lim_{k \to \infty} F(r^{k+1} z) / F(r^k z) = r < 1$. □

Substituting (3.17)–(3.19) into Eqs. (3.11)–(3.13), we have the generating functions

\[
\begin{align*}
\phi_1(x, z) &= \frac{S_1(x) - S_1(\tilde{\rho} + pz)}{x - (\tilde{\rho} + pz)} \frac{px(1 - z)\theta}{\theta S_1(\tilde{\rho} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{\rho} + pz)] + z\tilde{\theta} S_3(\tilde{\rho} + pz) - z} \phi_0(z), \\
\phi_2(x, z) &= \frac{S_2(x) - S_2(\tilde{\rho} + pz)}{px(1 - z)\theta S_1(\tilde{\rho} + pz)} \frac{z\tilde{\theta} S_3(\tilde{\rho} + pz)}{x - (\tilde{\rho} + pz)} \phi_0(z), \\
\phi_3(x, z) &= \frac{S_3(x) - S_3(\tilde{\rho} + pz)}{px(1 - z)\tilde{\theta}} \frac{z\tilde{\theta} S_3(\tilde{\rho} + pz)}{x - (\tilde{\rho} + pz)} \phi_0(z).
\end{align*}
\]

The normalizing condition $\phi_0(1) + \phi_1(1, 1) + \phi_2(1, 1) + \phi_3(1, 1) = 1$ allows us to find out the constant $\phi_0(1) = 1 - \rho$ and therefore $\phi_0(0)$.

We summarize the above results in the following theorem.

**Theorem 3.4.** The stationary distribution of the Markov chain $\{X_m, m \in \mathbb{N}\}$ is ergodic if and only if $\rho < 1$ and has the following generating functions:

\[
\phi_0(z) = (1 - \rho) \prod_{k=1}^{\infty} G(r^k z),
\]

\[
\prod_{k=1}^{\infty} G(r^k z).
\]
The probability generating function of the system size is given by
\[ \phi_0(x) = \frac{S_1(x) - S_1(\rho + pz)}{x - (\rho + pz)} \theta S_1(\rho + pz)\{\bar{\alpha} + \alpha S_2(\rho + pz)\} + z\theta S_3(\rho + pz) - z \]

The marginal generating function of the number of customers in the orbit when the server is busy is given by
\[ \phi_0(x) = \frac{S_2(x) - S_2(\rho + pz)}{x - (\rho + pz)} \theta S_1(\rho + pz)\{\bar{\alpha} + \alpha S_2(\rho + pz)\} + z\theta S_3(\rho + pz) - z \]

The mean orbit and system size is given by
\[ \phi_0(x) = \frac{S_3(x) - S_3(\rho + pz)}{x - (\rho + pz)} \theta S_1(\rho + pz)\{\bar{\alpha} + \alpha S_2(\rho + pz)\} + z\theta S_3(\rho + pz) - z \]

where
\[ G(z) = \frac{\theta S_1(\rho + pz)\bar{\alpha} + \alpha S_2(\rho + pz)}{(\rho + pz)\{\theta S_1(\rho + pz)\bar{\alpha} + \alpha S_2(\rho + pz)\} + z\theta S_3(\rho + pz) - z} \]

Corollary 3.5. In the steady state, we have
1. The marginal generating function of the number of customers in the orbit when the server is idle is given by \( \phi_0(z) \).
2. The marginal generating function of the number of customers in the orbit when the server is busy is given by
\[ \phi_1(z) + \phi_2(z) = \frac{\theta[1 - S_1(\rho + pz)] + \alpha \theta S_1(\rho + pz)[1 - S_2(\rho + pz)]}{\theta S_1(\rho + pz)\{\bar{\alpha} + \alpha S_2(\rho + pz)\} + z\theta S_3(\rho + pz) - z} \phi_0(z). \]
3. The probability generating function of the number of customers in the orbit when the server is down is given by
\[ \phi_3(z) = \frac{\bar{\theta}z[1 - S_3(\rho + pz)]}{\theta S_1(\rho + pz)\{\bar{\alpha} + \alpha S_2(\rho + pz)\} + z\theta S_3(\rho + pz) - z} \phi_0(z). \]
4. The probability generating function of the orbit size is given by
\[ \psi(z) = \phi_0(z) + \phi_1(z) + \phi_2(z) + \phi_3(z) = \frac{\theta(1 - z)}{\theta S_1(\rho + pz)\{\bar{\alpha} + \alpha S_2(\rho + pz)\} + z\theta S_3(\rho + pz) - z} \phi_0(z). \]
5. The probability generating function of the system size is given by
\[ \Phi(z) = \phi_0(z) + z\phi_1(z) + z\phi_2(z) + \phi_3(z) = \frac{(1 - z)\theta S_1(\rho + pz)\{\bar{\alpha} + \alpha S_2(\rho + pz)\}}{\theta S_1(\rho + pz)\{\bar{\alpha} + \alpha S_2(\rho + pz)\} + z\theta S_3(\rho + pz) - z} \phi_0(z). \]

Corollary 3.6. If \( \rho < 1 \), then we have
1. The stationary distribution of the server state is given by
\[ \phi_0(1) = 1 - \rho, \quad \phi_1(1) = \rho_1, \quad \phi_2(1) = \rho_2, \quad \phi_3(1) = \rho_3. \]
2. The mean orbit and system size is given by
\[ E[N] = \frac{p^2(\beta_{1,2} + \alpha \beta_{2,2}) + \rho_1 \rho_2 + \rho_3}{2(1 - \rho)} + \sum_{k=1}^{\infty} G'(r^k) r^k, \]
\[ E[L] = (\rho_1 + \rho_2) + E[N]. \]
3. The mean time a customer spends in the system is given by
\[ W = \frac{E[L]}{p}. \]

Remark 3.7. It can be observed the relation \( \Phi(z) = \psi(z)S_1(\rho + pz)[\bar{\alpha} + \alpha S_2(\rho + pz)] \), and as a consequence we find the formula
\[ \phi^{(n)}(1) = \sum_{m=0}^{n} C_n^m p^m (\beta_{1,m} + \alpha \beta_{2,m}) \psi^{(n-m)}(1), \quad n \geq 1 \]
where \( \Phi^{(n)}(1) \) and \( \Psi^{(n)}(1) \) are the \( n \)th factorial moments for the distribution of the random variables \( L \) and \( N \) respectively.

**Remark 3.8.** The stationary distribution of the server state

\[
\phi_0(1) = 1 - \rho, \quad \phi_1(1, 1) = \rho_1, \quad \phi_2(1, 1) = \rho_2, \quad \phi_3(1, 1) = \rho_3
\]
depends on the optional probability, the service and repair times distributions only through their means and is independent of the inter-retrial times parameter \( r \).

**Remark 3.9.** To approximate numerically the previous performance measures, it is necessary to estimate the series \( \sum_{k=1}^{\infty} \frac{G'(r^k)}{G(r^k)} r^k \). It is easy to see the relation

\[
\sum_{k=1}^{\infty} \frac{G'(r^k)}{G(r^k)} r^k = \frac{1}{1-r} \sum_{k=1}^{\infty} \frac{G'(r^k)}{G(r^k)} (r^k - r^{k+1}).
\]

Consequently, the former series can be approximated by an integral as follows:

\[
\sum_{k=1}^{\infty} \frac{G'(r^k)}{G(r^k)} r^k \approx \frac{r}{1-r} \int_0^1 \frac{G'(z)}{G(z)} dz = \frac{r \ln G(1)}{1-r} = \frac{r \ln \left[ p + \theta(1 - \rho_1 - \rho_2 - \rho_3) \right]}{1-r}
\]

if \( r \) is close to 1.

\[
\sum_{k=1}^{\infty} \frac{G'(r^k)}{G(r^k)} r^k \approx \frac{r}{1-r} \sum_{k=1}^{\infty} \frac{G'(r^k)}{G(r^k)} (r^k - r^{k+1}) + \frac{1}{1-r} \int_0^{n_0(\epsilon)+1} \frac{G'(z)}{G(z)} dz
\]

\[
= \sum_{k=1}^{n_0(\epsilon)} \frac{G'(r^k)}{G(r^k)} r^k + \frac{\ln G(r^{n_0(\epsilon)+1})}{1-r}
\]

where, for each \( \epsilon > 0, n_0(\epsilon) \) is chosen such that \( r^{n_0(\epsilon)+1} < \epsilon \).

The approximation (ii) is applicable for any \( r \in [0, 1) \), although it is complex to obtain the functions \( G(z) \) and \( G'(z) \).

**Remark 3.10 (Two Special Models).** One case is a queue with random service discipline and the other is a system with reliable server.

(a) When \( r = 0 \), \( \Phi(z) \) has the expression

\[
\phi(z) = \frac{(1-\rho)(1-z)\theta S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)]}{\theta S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)] + \bar{\beta} \bar{S}_3(\bar{p} + pz) - z},
\]

which is the probability generating function of the system size in the model Geo/G/1/\( \infty \) with starting failures and 2nd optional service. This is not surprising because as \( r = 0 \), the blocked customers try to get service at every slot boundary. This system is equivalent to the model Geo/G/1/\( \infty \) with starting failures, 2nd optional service and random service discipline, since the system size distribution does not depend on the service discipline.

(b) When \( \theta = 1 \), our model becomes a discrete-time Geo/G/1 retrial queue with 2nd optional service and reliable server. Therefore the generating functions of Theorem 3.4 are reduced to

\[
\phi_0(z) = (1 - \rho_1 - \rho_2) \prod_{k=1}^{\infty} \frac{G(r^k)}{G(r^k)},
\]

\[
\phi_1(x, z) = \frac{S_1(x) - S_1(\bar{p} + pz)}{x - (\bar{p} + pz)} \frac{px(1-z)}{S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)] - z} \phi_0(z),
\]

\[
\phi_2(x, z) = \frac{S_2(x) - S_2(\bar{p} + pz)}{x - (\bar{p} + pz)} \frac{\alpha px(1-z)}{S_1(\bar{p} + pz)[\bar{\alpha} + \alpha S_2(\bar{p} + pz)] - z} \phi_0(z),
\]

where \( \phi_0(z) \) and \( \phi_1(x, z) \) are the nth factorial moments for the distribution of the random variables \( L \) and \( N \) respectively.
where
\[
G(z) = \frac{S_1(\tilde{p} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{p} + pz)] - z(\tilde{p} + pz)}{(\tilde{p} + pz)[S_1(\tilde{p} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{p} + pz)] - z]} \tilde{p},
\]
which coincide with the generating functions of Theorem 1 in [20].

4. Stochastic decomposition laws

In this section we investigate the stochastic decomposition property of the system size distribution. The stochastic decomposition law for retrial queues has been studied by Yang and Templeton [30], and later by Artalejo and Falin [31].

In our case the probability generating function of the system size can be written as
\[
(\text{a}) \quad \Phi(z) = \frac{(1 - \rho_1 - \rho_2)(1 - z)S_1(\tilde{p} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{p} + pz)] - z(\tilde{p} + pz)}{S_1(\tilde{p} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{p} + pz)] - z} \phi_0(1) + \phi_3(1, z),
\]
whence
\[
(\text{b}) \quad \Phi(z) = \frac{(1 - \rho_1 - \rho_2)(1 - z)\theta S_1(\tilde{p} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{p} + pz)] - z \theta S_3(\tilde{p} + pz)}{\theta S_1(\tilde{p} + pz)[\tilde{\alpha} + \alpha S_2(\tilde{p} + pz)] - z} \phi_0(1).
\]
whose meaning will be explained in the following decomposition laws.

**Theorem 4.1.** (a) The total number \( L \) of customers in the system under study can be expressed as the sum of two independent random variables, one of which is the total number \( L' \) of customers in the Geo/G/1/∞ queueing system with 2nd optional service and the other one is the number \( M' \) of repeated customers given that the server is idle or down. That is, \( L = L' + M' \).

(b) The total number \( L \) of customers in the system under study can be expressed as the sum of two independent random variables, one of which is the total number \( L'' \) of customers in the Geo/G/1/∞ queueing system with starting failures and 2nd optional service and the other one is the number \( M'' \) of repeated customers given that the server is idle. That is, \( L = L'' + M'' \).

**Proof.** Firstly, it can be shown that in the above expression (a) the first fraction is the probability generating function of the number of customers in a Geo/G/1/∞ queue with second optional service (see [20]) and the second fraction is the probability generating function of the number of repeated customers given that the server is idle or down. In terms of distributions, this property means that \( L = L' + M' \).

In the same way, in the expression (b) the first fraction corresponds to the probability generating function of the number of customers in the Geo/G/1/∞ queue with second optional service and starting failures, and the second fraction corresponds to the probability generating function of the number of repeated customers given that the server is idle. That is, \( L = L'' + M'' \).  

**Theorem 4.2.** The following inequalities hold:
\[
2(1 - \rho_1 - \rho_2 - \pi_{0,0}) \leq \sum_{j=0}^{\infty} |P[L = j] - P[L' = j]| \leq 2 \frac{1 - \rho_1 - \rho_2 - \pi_{0,0}}{1 - \rho_1 - \rho_2},
\]
\[
2(1 - \rho_1 - \rho_2 - \rho_3 - \pi_{0,0}) \leq \sum_{j=0}^{\infty} |P[L = j] - P[L'' = j]| \leq 2 \frac{1 - \rho_1 - \rho_2 - \rho_3 - \pi_{0,0}}{1 - \rho_1 - \rho_2 - \rho_3}.
\]

**Proof.** The proof is based on the stochastic decomposition laws of the total number of customers in the system which are obtained in Theorem 4.1 (see [29]). In our case, we have
\[
P[L = j] = \sum_{k=0}^{j} P[L' = k]P[M' = j - k].
\]
Based on the previous results, the following can be obtained easily:
\[
|P[L = j] - P[L' = j]| \leq (1 - \delta_{0,j}) \sum_{k=0}^{j-1} P[L' = k]P[M' = j - k] + P[L' = j](1 - P[M' = 0]).
\]
Fig. 1, we will show various graphics to study the influence of the parameters (the retrial rate, condition, one finds the value \( \square \) here.

On the other hand, using the inequality \( |a - b| \geq a - b \), we have

\[
\sum_{j=0}^{\infty} |P[L = j] - P[L' = j]| \leq \sum_{j=0}^{\infty} P[L = j] + (1 - 2P[M' = 0]) \sum_{j=0}^{\infty} P[L' = j]
\]

\[
= 2(1 - 2P[M' = 0]) = 2 \frac{\phi_0(1) + \phi_3(1) - \pi_{0,0}}{\phi_0(1) + \phi_3(1)}.
\]

Summing over all the states, we get the upper bound:

\[
\sum_{j=0}^{\infty} |P[L = 0] - P[L' = 0]| = \sum_{j=1}^{\infty} (P[L = j] - P[L' = j])
\]

\[
= P[L' = 0](1 - P[M' = 0]) + 1 - P[L = 0] - 1 + P[L' = 0]
\]

\[
= 2P[L' = 0](1 - P[M' = 0]) = 2(\phi_0(1) + \phi_3(1) - \pi_{0,0}).
\]

This completes the proof of the first inequality. The second result can be proved in the same way, so we omit it here. \( \square \)

As a comment, we observe that the distance \( \sum_{j=0}^{\infty} |P[L = j] - P[L' = j]| \) between the distributions of the variables \( L \) and \( L' \) decreases as \( (\theta, r) \) approaches \((1,0)\); similarly the distance \( \sum_{j=0}^{\infty} |P[L = j] - P[L'' = j]| \) between the distributions of the variables \( L \) and \( L'' \) diminishes when \( r \) goes towards \( 0 \). We also comment that the interest of the former theorem is to provide lower and upper estimates for the distance between these distributions.

5. Numerical examples

In this section, we present two numerical examples to study the effect of the system parameters on the mean orbit size. To this end, it is assumed that service times of the 1st (essential) and the 2nd (optional) services and repair times follow a Negative Binomial with generating functions

\[ S_1(x) = \left( \frac{x}{2-x} \right)^{n_1}, \quad S_2(x) = \left( \frac{x}{2-x} \right)^{n_2} \quad \text{and} \quad S_3(x) = \left( \frac{2x}{3-x} \right)^{n_3}, \]

respectively. Let us note that service times of the 1st (essential) and the 2nd (optional) services and repair times are the sum of \( n_1, n_2 \) and \( n_3 \) independent random variables geometrically distributed with means \( 2 \) and \( 1.5 \), respectively. We will concentrate our attention on the mean number of customers in the orbit. We will also remark that the mean orbit size \( E[N] \) is plotted versus \( \alpha \), because this is the most specific parameter of our model. From the ergodicity condition, one finds the value

\[ \alpha^*(p, \theta, \beta_{1,1}, \beta_{2,1}, \beta_{3,1}) = \frac{1 - p\beta_{1,1} - \frac{\theta}{\bar{\theta}} p\beta_{3,1}}{p\beta_{2,1}} \]

such that the system is stable if and only if \( \alpha < \alpha^* \). Therefore, the domain of the functions, where graphics are represented, will be \([0, \alpha^*]\).

In the following Fig. 1, we will show various graphics to study the influence of the parameters (the retrial rate, the probability that a customer turns on the server successfully, the mean service time of the second service, and the mean repair time) on several performance characteristics. From different choices of the parameters, the curves show that \( E[N] \) is increasing as a function of \( \alpha \), as we expected. In addition, the mean number of repeated customers diverges when the parameter \( \alpha \) approximates to \( \alpha^* \). These graphics corroborate that the expectation \( E[N] \) increases with increasing values of \( \beta_{2,1}, \beta_{3,1} \) and \( r \), with decreasing values of \( \theta \).

Finally, as is to be expected, we note that the expectation \( E[N] \) in the Geo/G/1 retrial queue with 2nd optional service (which is obtained for \( \theta = 1 \)) provides us a lower bound for the corresponding mean value in our queueing
Fig. 1. The mean orbit size vs. $\alpha$.

The authors acknowledge two anonymous reviewers for their comments which have definitely helped to improve the presentation of this paper.

References


