Available online at www.sciencedirect.com
ScienceDirect
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
Journal of Computational and Applied Mathematics 223 (2009) 15–26
www.elsevier.com/locate/cam

The $M/G/1$ retrial queue: New descriptors of the customer's behavior

J. Amador[a], J.R. Artalejo[b,*]

[a] School of Statistics, Complutense University of Madrid, 28040 Madrid, Spain
[b] Faculty of Mathematics, Complutense University of Madrid, 28040 Madrid, Spain

Received 6 September 2007; received in revised form 13 December 2007

Abstract

We consider queuing systems where customers are not allowed to queue; instead of that they make repeated attempts, or retrials, in order to enter service after some time. The performance of telephone systems and communication networks modelled as retrial queues differs from standard waiting lines because typically the retrial group is an *invisible queue* which cannot be observed. As a result, the original flow of primary arrivals and the flow of repeated attempts become undistinguished. Our aim in this paper is to consider some aspects of this problem. Thus, we focus on the main retrial model of $M/G/1$ type and investigate the distribution of the successful and blocked events made by the primary customers and the retrial customers.

Keywords: Queuing; Retrials; Blocked and successful events; *PH* distribution; Telephone systems; Communication networks

1. Introduction

In this paper, we deal with a branch of the queuing theory, retrial queues, which is characterized by the following feature: a customer who cannot receive service (due to finite capacity of the system, balking, impatience, etc.) leaves the service area but after some random time returns to the system again. As a consequence, repeated attempts for service from the retrial pool of unsatisfied customers are superimposed on the ordinary stream of arrivals of first attempts.

We may find queues with returning customers in our daily activities as well as in many telephone and communication systems. The following examples motivate the interest of retrial queues in telephone systems and local computer networks.

Example 1. *Telephone systems*

Everybody has experienced that a telephone subscriber who obtains a busy signal repeats the call until the required connection is made. As a result, the flow of calls circulating in a telephone network consists of two parts: the flow of primary calls, which reflects the real wishes of the telephone subscribers, and the flow of repeated calls, which

* Corresponding author.
E-mail address: jesus_artalejo@mat.ucm.es (J.R. Artalejo).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2007.12.016

is the consequence of the lack of success of previous attempts. This consideration brings into focus the need of the retrial queues as a proper modelling of customer behavior in modern telephone systems (cellular networks, auto-repeat facilities, repeat-last-number, etc.).

Example 2. *Random access protocols in communication networks*

Consider a communication line with slotted time which is shared by several stations. If two or more stations are transmitting packets simultaneously then a collision takes place. As a result, all the involved packets are destroyed and must be retransmitted. The stations would try to retransmit in the nearest slot but then a collision occurs with certainty. To avoid this, each station, independently of the other stations involved in the conflict, may either retransmit the packet with probability p or delay actions until the next slot with probability $1 - p$. In other words, each station introduces a random delay before the next attempt to transmit the packet.

Apart from their practical applications, the retrial queues raise interesting mathematical problems. Because of the retrial feature, the underlying stochastic process that represents the state of the system is not space-homogeneous. This creates severe analytical problems and allows explicit formulae only in a few special cases ($M/H_2/1$ retrial queue, $M/M/c$ retrial queue with $c \leq 2$). In contrast, the retrial literature is rich in approximating and numerical methods (see [4–6] and the references therein).

Our main goal in this paper is to investigate the distribution of the following performance descriptors of the customer's behavior: (i) the successful retrials, (ii) the successful arrivals, and (iii) the blocked retrials. We will show that the number of blocked primary arrivals amounts to the number of successful retrials referred to a busy period. Hence, our three measures provide a full description of what is relevant in practice in order to distinguish between primary arrivals and repeated attempts behaviors. The knowledge of the distribution of these descriptors seems of interest in its own right. We also refer the reader to Section 7, where we will illustrate the applicability in optimal design problems. Our attention is paid to methods for the computational analysis of the performance measures under study. To this end, we propose direct methods for the computation of the probability mass functions instead of using an alternative approach based on generating functions. The main advantage of the direct approach is in avoiding the numerical inversion of the generating functions.

As related work we mention the recent work of the authors [2] who introduced these descriptors for the main multiserver model of $M/M/c$ type. Although we study the same descriptors, the approach in this paper is different. The $M/M/c$ retrial queue is a Markovian model and, consequently, it is investigated at the epochs when any transition (i.e., arrival, departure, successful retrial) occurs. In contrast, the $M/G/1$ retrial queue is a semi-Markovian model which must be studied at the service completion epochs. As a result, the analysis of the $M/G/1$ queue is more involved. In particular, the study of the number of blocked retrials becomes far harder because an arbitrary number of blocked primary arrivals occurring during the current service time influence the number of retrials. We also remark the paper by Artalejo and Falin [3] where two structural characteristics of the orbit, namely, the orbit busy period and the orbit idle period, were investigated. To illustrate the active role of the retrial queues over the last few years, we mention a recent paper published in this journal [21] as well as some other recent publications [1,8–12,14–16,18–20].

In the next section we describe the queuing model under consideration; this is followed by Section 3 where we define our performance descriptors and state some basic relationships among them. In subsequent Sections 4–6, we deal with the computational analysis of the three proposed descriptors. Simple probabilistic arguments lead to an efficient recursive scheme for the exact computation of the probability mass function of the number of successful retrials. The equations governing the dynamics of the number of successful arrivals and the number of blocked retrials become more involved. As a result, we propose some approximating assumptions (truncation of the retrial group, use of phase type (PH) distributions, reallocation of arrivals). In Section 7 we present some numerical results including an optimal design problem. Finally, in Section 8, we suggest possible directions for further studies.

2. The queuing model

We next describe the main single-server queuing system of $M/G/1$ type. Primary customers arrive according to a Poisson process of rate λ. If a primary customer finds the server free, then he/she automatically receives service. The service times are general with probability distribution function $B(x)$ ($B(0) = 0$), Laplace–Stieltjes transform $\beta(s)$ and moments β_k, for $k \geq 1$. On the other hand, any arriving customer who finds the server busy leaves the service area and joins a retrial group called *orbit*. The policy of access from the orbit to the server is governed by an exponential law with rate $j\mu$, given that the number of customers in orbit is $j \geq 0$. We assume that the flow of primary arrivals, intervals between repeated attempts and service times are mutually independent.

Note that the state of the system at time t can be described by the process $X(t) = (C(t), N(t), \xi(t))$, where $C(t)$ is the state of the server (free/busy) and $N(t)$ is the number of customers in orbit at time t. If $C(t) = 1$, then $\xi(t)$ denotes the time from the last service starting before time t. In what follows, we neglect the elapsed time $\xi(t)$, so the system state reduces to $S = \{0, 1\} \times \mathbb{N}$. Let $\rho = \lambda\beta_1$ be the utilization factor. If $\rho < 1$ the system is stable and, in particular, the busy period is finite with probability 1. Moreover, if $\rho < 1$ then the limiting probabilities $p_{ij} = \lim_{t\to\infty} P\{C(t) = i, N(t) = j\}$ exist and are positive.

3. The performance descriptors

We now define our performance descriptors of the customer's behavior: R^s is the number of successful retrials during a busy period, A^s is the number of successful arrivals during a busy period and R^b is the number of blocked retrials during a busy period. The three measures are referred to a busy period which is defined as the first passage time to (0, 0), given that the initial state is (1, 0).

Note that the number of times per busy period that process X crosses up from $j-1$ to j customers in orbit (i.e., a blocked arrival occurs) is equal to the number of down crosses from j to $j-1$ (i.e., a successful retrial occurs). Thus, we get

$$R^s = A^b, \tag{1}$$

where A^b denotes the number of blocked arrivals during a busy period. Since the two descriptors are coincident, we only need to study one of them, we say R^s.

We also notice the two basic relationships:

$$I = 1 + A^s + A^b, \tag{2}$$

$$\sum_{i=1}^{I} R_i = R^s + R^b, \tag{3}$$

where I denotes the number of customers served during a busy period and R_i is the number of repeated attempts made by the i-th arrival until it gets service.

The relationships (1)–(3) also hold for the $M/M/c$ retrial queue [2]. Then, it can be proved that

$$E[R^s] = E[A^b],$$
$$E[I] = 1 + E[A^s] + E[A^b],$$
$$E[I]E[R] = E[R^s] + E[R^b],$$

where R denotes the number of repeated attempts made by an arbitrary retrial customer.

It is known [4,7] that

$$E[I] = p_{00}^{-1} = \frac{1}{1-\rho} \exp\left\{\frac{\lambda}{\mu}\int_0^1 \frac{1-\beta(\lambda-\lambda u)}{\beta(\lambda-\lambda u)-u}\mathrm{d}u\right\}, \tag{4}$$

$$E[R] = \frac{\rho}{1-\rho} + \frac{\lambda\beta_2\mu}{2(1-\rho)}. \tag{5}$$

Thus, we easily find that

$$E[R^s] = E[A^b] = p_{00}^{-1}\rho, \tag{6}$$

$$E[A^s] = p_{00}^{-1}(1-\rho) - 1, \tag{7}$$

$$E[R^b] = p_{00}^{-1}\left(E[R] - \rho\right). \tag{8}$$

4. The number of successful retrials

In this section, we derive recursive equations for the exact computation of the probability mass function $P\{R^s = r\}$, for $r \geq 0$. Let $x_i^s(r)$ be the probability of having exactly $r \geq 0$ successful retrials during the remaining busy period, given that a service time has just been completed leaving behind i customers in orbit, for $0 \leq i \leq r$. We notice that $x_0^s(r) = \delta_{r0}$, for $r \geq 0$, where δ_{r0} denotes Kronecker's function.

Then, the probability distribution of R^s satisfies that

$$P\{R^s = r\} = \sum_{i=0}^{r} c_i x_i^s(r), \quad r \geq 0, \tag{9}$$

where $c_k = \int_0^\infty \mathrm{e}^{-\lambda x}(\lambda x)^k/k!\mathrm{d}B(x)$, for $k \geq 0$, is the probability that k arrivals occur during a service time.

In the next theorem, we derive equations for the probabilities $x_i^s(r)$.

Theorem 1. *For each fixed $r \geq 1$, the probabilities $x_i^s(r)$ can be written in matrix form as*

$$\mathbf{M}_r \mathbf{x}_r^s = \mathbf{B}_r \widetilde{\mathbf{x}}_r, \tag{10}$$

where $\mathbf{x}_r^s = \left(x_1^s(r), \ldots, x_r^s(r)\right)'$, $\widetilde{\mathbf{x}}_r = \left(\delta_{r1}, \mathbf{x}_{r-1}^s\right)'$ ($\widetilde{\mathbf{x}}_1 \equiv 1$), and $\mathbf{M}_r = \left(m_{ij}\right)$ and $\mathbf{B}_r = \left(b_{ij}\right)$ are square matrices of order r with elements

$$m_{ij} = \begin{cases} 0, & \text{if } 1 \leq j < i \leq r, \\ \lambda(1-c_0) + i\mu, & \text{if } j = i, \\ -\lambda c_{j-i}, & \text{if } 1 \leq i < j \leq r, \end{cases} \qquad b_{ij} = \begin{cases} 0, & \text{if } 1 \leq j < i \leq r, \\ i\mu c_{j-i}, & \text{if } 1 \leq i \leq j \leq r. \end{cases}$$

Proof. We analyze the motion between two successive service completion epochs in order to find the dynamics of the probabilities $x_i^s(r)$. This gives

$$x_i^s(r) = \frac{\lambda}{\lambda + i\mu} \sum_{j=i}^{r} c_{j-i} x_j^s(r) + \frac{i\mu}{\lambda + i\mu} \sum_{j=i-1}^{r-1} c_{j-i+1} x_j^s(r-1), \quad 1 \leq i \leq r. \tag{11}$$

Eq. (11) is derived by noting that the previous departure left i customers in orbit. Then, the next customer getting service may be a primary arrival (with probability $\lambda(\lambda + i\mu)^{-1}$) or may come from the orbit (with probability $i\mu(\lambda + i\mu)^{-1}$). In either case, we must record the number of arrivals occurring during the subsequent service time so that, at its completion, the accumulated number of customers who visited the orbit never reaches $r + 1$.

For each fixed $r \geq 1$, we put (11) in matrix form which leads to (10). $\square$

Since the system (10) is upper triangular, it is suitable for recursively getting the probabilities $x_i^s(r)$ and, consequently, $P\{R^s = r\}$ is determined from (9).

5. The number of successful arrivals

We now turn our attention to the study of the number of successful primary arrivals. We approximate the $M/G/1$ retrial queue with infinite orbit by the truncated model with orbit capacity $K \geq 1$. Blocked customers finding the state $(1, K)$ upon arrival are lost. The necessity of dealing with a truncated system will be explained at the end of this section.

Let $x_i^s(a)$ be the probability of having $a \geq 0$ successful arrivals during the remaining busy period, given that a service time has been completed leaving behind i customers in orbit, for $0 \leq i \leq K$. We notice that $x_0^s(a) = \delta_{a0}$. Now, we have

$$P\{A^s = a\} = \sum_{i=0}^{K-1} c_i x_i^s(a) + \left(1 - \sum_{i=0}^{K-1} c_i\right) x_K^s(a), \quad a \geq 0. \tag{12}$$

Theorem 2. *For each fixed $a \geq 0$, the probabilities $x_i^s(a)$ satisfy the following system:*

$$\mathbf{M}\mathbf{x}_0^s = \mathbf{b}_0^s, \tag{13}$$

$$\mathbf{M}\mathbf{x}_a^s = \mathbf{B}\mathbf{x}_{a-1}^s, \quad a \geq 1, \tag{14}$$

where $\mathbf{x}_a^s = \left(x_1^s(a), \ldots, x_K^s(a)\right)'$, $\mathbf{b}_0^s = (\mu c_0, 0, \ldots, 0)'$, *and* $\mathbf{M} = \left(m_{ij}\right)$ *and* $\mathbf{B} = \left(b_{ij}\right)$ *are square matrices of order K with elements*

$$m_{ij} = \begin{cases} 0, & \text{if } 3 \leq i \leq K,\ 1 \leq j \leq i-2, \\ (\lambda + i\mu) - i\mu c_1, & \text{if } 1 \leq i \leq K-1,\ j = i, \\ (\lambda + K\mu) - K\mu(1 - c_0), & \text{if } i = j = K, \\ -i\mu c_0, & \text{if } 2 \leq i \leq K,\ j = i-1, \\ -i\mu c_{j-i+1}, & \text{if } 1 \leq i \leq K-2,\ i+1 \leq j \leq K-1, \\ -i\mu\left(1 - \sum_{k=0}^{K-i} c_k\right), & \text{if } 1 \leq i \leq K-1,\ j = K. \end{cases}$$

$$b_{ij} = \begin{cases} 0, & \text{if } 1 \leq j < i \leq K, \\ \lambda c_{j-i}, & \text{if } 1 \leq i \leq j \leq K-1, \\ \lambda\left(1 - (1 - \delta_{iK}) \sum_{k=0}^{K-i-1} c_k\right), & \text{if } 1 \leq i \leq K,\ j = K. \end{cases}$$

Proof. First of all, we notice that

$$x_i^s(a) = (1 - \delta_{a0}) \frac{\lambda}{\lambda + i\mu} \left((1 - \delta_{iK}) \sum_{j=i}^{K-1} c_{j-i} x_j^s(a-1) + \left(1 - (1 - \delta_{iK}) \sum_{j=0}^{K-i-1} c_j\right) x_K^s(a-1) \right)$$
$$+ \frac{i\mu}{\lambda + i\mu} \left(\sum_{j=i-1}^{K-1} c_{j-i+1} x_j^s(a) + \left(1 - \sum_{j=0}^{K-i} c_j\right) x_K^s(a) \right), \quad 1 \leq i \leq K,\ a \geq 0. \tag{15}$$

With the help of Kronecker's function δ_{ab}, we have written the above compact formula, but it is convenient to derive the expression by thinking in three different cases: $(a = 0)$, $(a \geq 1, 1 \leq i < K)$ and $(a \geq 1, i = K)$. If the primary arrival wins the competition to occupy the free server, then the index a decreases one unit. Once more, we must update the orbit state by counting the number of arrivals during the service time in progress.

After rearrangement, the system (15) can be written in matrix form as claimed in (13) and (14). □

In the light of formulas (12) and (15) we can understand the need of considering a truncated model. Otherwise, the sums would turn into infinite series. As a result, for each fixed a, we would have an infinite system of Eq. (15) with a matrix $\mathbf{M}$ of level dependent upper Hessenberg type. Unfortunately, such a system of equations has no known solution. Of course, it remains to specify how to choose the truncation threshold K. This discussion is postponed to Section 7.

6. The number of blocked retrials

The number of blocked retrials taking place during a given service time depends on the arbitrary number of blocked primary arrivals occurring during the service time in progress. Thus, it seems difficult, or even impossible, to obtain the exact distribution of R^b. In this section, we propose two methods for the computation of the probability mass function of R^b for the model with finite capacity $K \geq 1$. The aim of the first approximation is to present a tractable model for accurately representing the service times. To this end, we assume that service times follow a *PH* distribution [13]. Phase type distributions form a versatile family of probability distributions. The exponential, Erlang and hyperexponential distributions belong to this family. A remarkable property is that the set of *PH* distributions is dense in the set of all probability distributions over $[0,\infty)$. Moreover, sums and mixtures of independent *PH* random variables lead again *PH* distributions.

We assume a phase type representation $(\boldsymbol{\alpha}, \mathbf{T})$ of order s, where the row vector $\boldsymbol{\alpha}$ gives the initial phase distribution and matrix $\mathbf{T}$ governs the infinitesimal phase rates. We also define the column vector $\mathbf{t} = -\mathbf{T}\mathbf{e}(s)$, where $\mathbf{e}(r)$

denotes the column vector of dimension r consisting of 1's. For the use in the sequel, let $\mathbf{e}_j(r)$, $\mathbf{E}_{ij}(r)$ and $\mathbf{I}_r$ denote, respectively, the column vector of dimension r with 1 in the j-th position and 0 elsewhere, the matrix $\mathbf{e}_i(r)\mathbf{e}'_j(r)$ and an identity matrix of dimension r.

The state of the $M/PH/1/K$ queue with retrials is described by the Markov chain $Y(t) = (K(t), N(t))$, where $K(t)$ is the phase of the service in progress. Note that $K(t) = 0$ indicates that the server is idle. The infinitesimal generator, in partitioned form, is given by

$$\mathbf{Q} = \begin{pmatrix} \mathbf{A}_0^{(0)} & \mathbf{A}_0^{(+1)} & \mathbf{0}_{s+1} & \cdots & \mathbf{0}_{s+1} & \mathbf{0}_{s+1} \\ \mathbf{A}_1^{(-1)} & \mathbf{A}_1^{(0)} & \mathbf{A}_1^{(+1)} & \cdots & \mathbf{0}_{s+1} & \mathbf{0}_{s+1} \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{0}_{s+1} & \mathbf{0}_{s+1} & \mathbf{0}_{s+1} & \cdots & \mathbf{A}_{K-1}^{(0)} & \mathbf{A}_{K-1}^{(+1)} \\ \mathbf{0}_{s+1} & \mathbf{0}_{s+1} & \mathbf{0}_{s+1} & \cdots & \mathbf{A}_K^{(-1)} & \mathbf{A}_K^{(0)} \end{pmatrix}, \tag{16}$$

where $\mathbf{0}_{s+1}$ is a matrix of zeros of dimension $s + 1$, and the coefficient matrices appearing in (16) are given by

$$\mathbf{A}_j^{(-1)} = \begin{pmatrix} 0 & j\mu\boldsymbol{\alpha} \\ \mathbf{0}_{s\times(s+1)} & \end{pmatrix}, \quad 1 \le j \le K,$$
$$\mathbf{A}_j^{(0)} = \begin{pmatrix} -(\lambda + j\mu) & \lambda\boldsymbol{\alpha} \\ \mathbf{t} & -(1-\delta_{jK})\lambda\mathbf{I}_s + \mathbf{T} \end{pmatrix}, \quad 0 \le j \le K,$$
$$\mathbf{A}^{(+1)} = \mathbf{A}_j^{(+1)} = \lambda\left(\mathbf{I}_{s+1} - \mathbf{E}_{11}(s+1)\right), \quad 0 \le j \le K-1.$$

Firstly, we condition on the phase of the first arrival of the busy period. Then, we have

$$P\{R^b = r\} = \sum_{k=1}^{s} \alpha_k y_{k0}^b(r), \quad r \ge 0,$$

where $y_{kj}^b(r)$ is the probability of having r blocked retrials during the remaining busy period, given that the current system state is (k, j), for $0 \le k \le s$ and $0 \le j \le K$.

The unknowns $y_{kj}^b(r)$ satisfy the following.

Theorem 3. *For each fixed $r \ge 0$, the probabilities $y_{kj}^b(r)$ can be written in matrix form as*

$$\mathbf{P}\mathbf{y}_r^b = \mathbf{f}_r^b, \tag{17}$$

where $\mathbf{y}_r^b = \left(\mathbf{y}_0^b(r), \ldots, \mathbf{y}_K^b(r)\right)'$ and $\mathbf{y}_j^b(r) = \left(y_{0j}^b(r), \ldots, y_{sj}^b(r)\right)'$, for $0 \le j \le K$, $\mathbf{f}_0^b = -\mathbf{e}_1((s+1)(K+1))$ and $\mathbf{f}_r^b = -\mathbf{R}\mathbf{y}_{r-1}^b$, for $r \ge 1$, and

$$\begin{aligned} \mathbf{P} &= \mathbf{Q} - \mathbf{R} - \mathbf{F}, \\ \mathbf{R} &= \frac{\mu}{\lambda}\mathrm{Diag}(0, 1, \ldots, K) \otimes \mathbf{A}^{(+1)}, \\ \mathbf{F} &= \mathbf{E}_{11}((s+1)(K+1))\left(\mathbf{Q} + \mathbf{E}_{11}((s+1)(K+1))\right). \end{aligned} \tag{18}$$

Proof. The matrix $\mathbf{Q}$ describes the infinitesimal dynamics of the $M/PH/1/K$ queue with retrials including (i) the blocked arrivals (see $\mathbf{A}^{(+1)}$), (ii) the successful retrials (see $\mathbf{A}_j^{(-1)}$), (iii) the successful arrivals (see $\lambda\boldsymbol{\alpha}$ in $\mathbf{A}_j^{(0)}$), (iv) the service completions (see $\mathbf{t}$ in $\mathbf{A}_j^{(0)}$), and (v) the phase transitions (see $\mathbf{T}$ in $\mathbf{A}_j^{(0)}$). The blocked retrials do not cause a transition in the Markov chain, so their effect is not represented in the infinitesimal generator $\mathbf{Q}$. However, given the state (k, j) with $1 \le k \le s$, the blocked retrials influence the dynamics of the probability $y_{kj}^b(r)$ leading to the transition

$$(1-\delta_{r0})\frac{j\mu}{\lambda - \mathbf{T}_{kk} + j\mu} y_{kj}^b(r-1).$$

The effect of the above transition is reflected in the matrix $\mathbf{R}$ and in the vector $\mathbf{f}_r^b$. On the other hand, the effect of subtracting the matrix $\mathbf{F}$ in (18) is to introduce $y_{00}^b(r) = \delta_{r0}$ as the first unknown of the system of equations. From these observations, formula (17) is easily derived. □

To solve the finite block tridiagonal systems in (17), we recommend using a block forward-elimination–backward-substitution algorithm [13]. However, other methods such as aggregate/disaggregate techniques and block Gaussian–Seidel iteration can also be employed. For the sake of ease, in what follows we present Eq. (17) in the particular cases where the service times follow hyperexponential and Erlang laws.

1. *Hyperexponential service times*

For the case of hyperexponential service times, that is, $B(x) = \sum_{k=1}^{s} p_k \left(1 - \mathrm{e}^{-\nu_k x}\right)$, for $x > 0$, $p_k > 0$, and $\sum_{k=1}^{s} p_k = 1$; the probabilities $P\{R^b = r\}$, for $r \geq 0$, can be calculated as

$$P\{R^b = r\} = \sum_{k=1}^{s} p_k y_{k0}^b(r), \quad r \geq 0,$$

where the unknowns $y_{kj}^b(r)$ satisfy the following system:

$$\begin{aligned}
y_{00}^b(r) &= \delta_{r0}, \\
y_{0j}^b(r) &= \frac{\lambda}{\lambda + j\mu} \sum_{k=1}^{s} p_k y_{kj}^b(r) + \frac{j\mu}{\lambda + j\mu} \sum_{k=1}^{s} p_k y_{k,j-1}^b(r), \quad 1 \leq j \leq K, \\
y_{kj}^b(r) &= \frac{\lambda}{\lambda + \nu_k + j\mu} y_{k,j+1}^b(r) + \frac{\nu_k}{\lambda + \nu_k + j\mu} y_{0j}^b(r) + (1 - \delta_{r0}) \frac{j\mu}{\lambda + \nu_k + j\mu} y_{kj}^b(r-1), \\
&\quad 1 \leq k \leq s,\ 0 \leq j \leq K-1, \\
y_{kK}^b(r) &= \frac{\nu_k}{\nu_k + K\mu} y_{0K}^b(r) + (1 - \delta_{r0}) \frac{K\mu}{\nu_k + K\mu} y_{kK}^b(r-1), \quad 1 \leq k \leq s.
\end{aligned}$$

2. *Erlang service times*

In this case, we have $B(x) = \int_0^x \frac{\nu^s}{(s-1)!} t^{s-1} \mathrm{e}^{-\nu t} \mathrm{d}t$, for $x > 0$ and $s \in \{1, 2, \ldots\}$. Then, we have

$$P\{R^b = r\} = y_{10}^b(r), \quad r \geq 0,$$

and

$$\begin{aligned}
y_{00}^b(r) &= \delta_{r0}, \\
y_{0j}^b(r) &= \frac{\lambda}{\lambda + j\mu} y_{1j}^b(r) + \frac{j\mu}{\lambda + j\mu} y_{1,j-1}^b(r), \quad 1 \leq j \leq K, \\
y_{kj}^b(r) &= \frac{\lambda}{\lambda + \nu + j\mu} y_{k,j+1}^b(r) + \frac{\nu}{\lambda + \nu + j\mu} y_{k+1,j}^b(r) + (1 - \delta_{r0}) \frac{j\mu}{\lambda + \nu + j\mu} y_{kj}^b(r-1), \\
&\quad 1 \leq k \leq s-1,\ 0 \leq j \leq K-1, \\
y_{kK}^b(r) &= \frac{\nu}{\nu + K\mu} y_{k+1,K}^b(r) + (1 - \delta_{r0}) \frac{K\mu}{\nu + K\mu} y_{kK}^b(r-1), \quad 1 \leq k \leq s-1, \\
y_{sj}^b(r) &= \frac{\lambda}{\lambda + \nu + j\mu} y_{s,j+1}^b(r) + \frac{\nu}{\lambda + \nu + j\mu} y_{0j}^b(r) + (1 - \delta_{r0}) \frac{j\mu}{\lambda + \nu + j\mu} y_{sj}^b(r-1), \\
&\quad 0 \leq j \leq K-1, \\
y_{sK}^b(r) &= \frac{\nu}{\nu + K\mu} y_{0K}^b(r) + (1 - \delta_{r0}) \frac{K\mu}{\nu + K\mu} y_{sK}^b(r-1).
\end{aligned}$$

As we mentioned earlier, the *PH* distribution provides a versatile representation of the service times which can be satisfactory for practical purposes. However, we may still wish to deal with any arbitrary service time, not necessarily of *PH* type. This can be done with the help of the following approximating approach.

We condition on the number of arrivals and retrials occurring during the first service time of the busy period. This gives

$$P\{R^b = r\} = \delta_{r0} c_0^{0,0} + (1-\delta_{K1}) \sum_{i=1}^{K-1} \sum_{k=0}^{r} c_0^{i,k} x_i^b(r-k) + \sum_{k=0}^{r} d_0^{K,k} x_K^b(r-k), \quad r \geq 0, \tag{19}$$

where $c_i^{j,k}$ is the probability that j primary arrivals and k retrials occur during a service time, given that immediately after the beginning of the service time the orbit size is i, for $0 \leq i \leq K$, $j \geq 0$ and $k \geq 0$. Then, the accumulative probabilities $d_i^{K-i,k}$ are defined by $d_i^{K-i,k} = \sum_{m=K-i}^{\infty} c_i^{m,k}$, for $0 \leq i \leq K$ and $k \geq 0$. The quantities $x_i^b(r)$ denote the probability of having r blocked retrials during the remaining busy period, given that a service time has been completed leaving i customers in orbit, for $0 \leq i \leq K$ and $r \geq 0$. For $i = 0$, we have $x_0^b(r) = \delta_{r0}$.

The dynamics of the probabilities $x_i^b(r)$ is summarized in the following result.

Theorem 4. *For each fixed $r \geq 0$, the probabilities $x_i^b(r)$ satisfy the following block tridiagonal system:*

$$\mathbf{N}_0 \mathbf{x}_0^b = \mathbf{b}_0^b, \tag{20}$$

$$\mathbf{N}_0 \mathbf{x}_r^b = \sum_{m=1}^{r} \mathbf{N}_m \mathbf{x}_{r-m}^b, \quad r \geq 1, \tag{21}$$

where $\mathbf{x}_r^b = \left(x_1^b(r), \ldots, x_K^b(r)\right)'$, for $r \geq 0$, $\mathbf{b}_0^b = \left(\mu \widehat{c}_0^{0,0}, 0, \ldots, 0\right)'$, and $\mathbf{N}_m = \left(n_{ij}^m\right)$ are square matrices of order K with elements

$$n_{ij}^0 = \begin{cases} 0, & \text{if } 3 \leq i \leq K,\ 1 \leq j \leq i-2, \\ (\lambda + i\mu) - \lambda \widehat{c}_i^{0,0} - i\mu \widehat{c}_{i-1}^{1,0}, & \text{if } 1 \leq i \leq K-1,\ j = i, \\ (\lambda + K\mu) - \lambda \widehat{d}_K^{0,0} - K\mu \widehat{d}_{K-1}^{1,0}, & \text{if } i = j = K, \\ -i\mu \widehat{c}_{i-1}^{0,0}, & \text{if } 2 \leq i \leq K,\ j = i-1, \\ -\lambda \widehat{c}_i^{j-i,0} - i\mu \widehat{c}_{i-1}^{j-i+1,0}, & \text{if } 1 \leq i \leq K-2,\ i+1 \leq j \leq K-1, \\ -\lambda \widehat{d}_i^{K-i,0} - i\mu \widehat{d}_{i-1}^{K-i+1,0}, & \text{if } 1 \leq i \leq K-1,\ j = K, \end{cases}$$

and, for $1 \leq m \leq r$,

$$n_{ij}^m = \begin{cases} 0, & \text{if } 3 \leq i \leq K,\ 1 \leq j \leq i-2, \\ \lambda \widehat{c}_i^{0,m} + i\mu \widehat{c}_{i-1}^{1,m}, & \text{if } 1 \leq i \leq K-1,\ j = i, \\ \lambda \widehat{d}_K^{0,m} + K\mu \widehat{d}_{K-1}^{1,m}, & \text{if } i = j = K, \\ i\mu \widehat{c}_{i-1}^{0,m}, & \text{if } 2 \leq i \leq K,\ j = i-1, \\ \lambda \widehat{c}_i^{j-i,m} + i\mu \widehat{c}_{i-1}^{j-i+1,m}, & \text{if } 1 \leq i \leq K-2,\ i+1 \leq j \leq K-1, \\ \lambda \widehat{d}_i^{K-i,m} + i\mu \widehat{d}_{i-1}^{K-i+1,m}, & \text{if } 1 \leq i \leq K-1,\ j = K. \end{cases}$$

The estimations $\widehat{c}_i^{j,k}$ and $\widehat{d}_i^{K-i,k}$ will be specified in the sequel.

Proof. Conditioning on the identity of the customer who occupies the server and counting the number of primary arrivals and retrials taking place during the service in progress, we find that

$$x_i^b(r) = \frac{\lambda}{\lambda + i\mu} \sum_{k=0}^{r} \left((1-\delta_{iK}) \sum_{j=i}^{K-1} c_i^{j-i,k} x_j^b(r-k) + d_i^{K-i,k} x_K^b(r-k) \right)$$
$$+ \frac{i\mu}{\lambda + i\mu} \sum_{k=0}^{r} \left(\sum_{j=i-1}^{K-1} c_{i-1}^{j-i+1,k} x_j^b(r-k) + d_{i-1}^{K-i+1,k} x_K^b(r-k) \right), \quad 1 \leq i \leq K,\ r \geq 0. \tag{22}$$

To specify the auxiliary quantities $c_i^{j,k}$ and $d_i^{K-i,k}$, suppose that the length of the service time is x and j primary customers arrive at epochs $x_1, \dots, x_j$. This implies that they may perform retrials during the remaining service time of length $x - x_1, \dots, x - x_j$, respectively. It seems very hard to manage these multidimensional constraints. Alternatively, we will assume that the j customers arrive at the mean point $x/2$. This simple approximating assumption affects on average only $\rho < 1$ customers per service. As a result of this reallocation, the constants $c_i^{j,k}$ are approximated by

$$\widehat{c}_i^{j,k} = \int_0^\infty \mathrm{e}^{-\lambda x} \frac{(\lambda x)^j}{j!} \sum_{m=0}^{k} \mathrm{e}^{-i\mu x} \frac{(i\mu x)^m}{m!} \mathrm{e}^{-\frac{j\mu x}{2}} \frac{(\frac{j\mu x}{2})^{k-m}}{(k-m)!} \mathrm{d}B(x).$$

Now, it is easy to derive that

$$\widehat{c}_i^{j,k} = \frac{\lambda^j \left(i\mu + \frac{j\mu}{2}\right)^k}{j!k!} \int_0^\infty \mathrm{e}^{-\left(\lambda + i\mu + \frac{j\mu}{2}\right)x} x^{j+k} \mathrm{d}B(x), \tag{23}$$

$$\begin{aligned} \widehat{d}_i^{K-i,k} = \frac{\left(\frac{(K+i)\mu}{2}\right)^k}{k!} & \left(\int_0^\infty \mathrm{e}^{-\frac{(K+i)\mu x}{2}} x^k \mathrm{d}B(x) \right. \\ & \left. - (1 - \delta_{iK}) \sum_{j=0}^{K-i-1} \frac{\lambda^j}{j!} \int_0^\infty \mathrm{e}^{-\left(\lambda + \frac{(K+i)\mu}{2}\right)x} x^{j+k} \mathrm{d}B(x) \right). \end{aligned} \tag{24}$$

By expressing (22) in matrix form, we obtain (20) and (21). □

Explicit expressions for approximated quantities (23) and (24) can be given for the most usual service time distributions. For example, in the case of exponential service times (i.e., $\beta_1 = \nu^{-1}$) formula (23) yields

$$\widehat{c}_i^{j,k} = \binom{j+k}{j} \frac{\lambda^j \nu \left(i\mu + \frac{j\mu}{2}\right)^k}{\left(\lambda + \nu + i\mu + \frac{j\mu}{2}\right)^{j+k+1}}.$$

The combination of formula (19) and the iterative solution of the systems (20) and (21) complete our second approach for the computation of the probabilities $P\{R^b = r\}$. For a comparison between the two approximating methods, we refer the reader to Tables 3 and 4 in Section 7.

7. Numerical results

In order to evaluate the performance of the descriptors under study, we next present some numerical experiments.

In Table 1, we display the expected value $E[R^s]$ for different choices of the service time distribution. To this end, we consider Erlang-3 (E_3), exponential (M) and hyperexponential (H_2) service times. The coefficient of variation of the hyperexponential law is 1.25 whereas β_1 has been normalized to be 1 in all cases. The traffic intensity ρ and the retrial rate μ take values 0.2, 0.4, 0.6 and 0.8, and 0.05, 0.5, 2.5, 25.0 and 100.0, respectively. Each cell is associated with a pair (ρ, μ) and gives the expectation for the three service time laws. An examination of the table reveals that $E[R^s]$ is an increasing function of ρ but it decreases as function of μ. We also notice that $E[R^s(H_2)] < E[R^s(M)] < E[R^s(E_3)]$.

We have produced the parallel tables (not reported here) for the descriptors $E[A^s]$ and $E[R^b]$. The behavior of $E[A^s]$ is similar to that shown in Table 1 for $E[R^s]$. However, the analysis of $E[R^b]$ is more complicated. The expectation $E[R^b]$ is also increasing with ρ but it exhibits a minimum as function of μ. On the other hand, the relationships among the expectations $E[R^b(H_2)]$, $E[R^b(M)]$ and $E[R^b(E_3)]$ now depend on the choice of the pair (ρ, μ). The expected values have been computed from formulas (4)–(8), where we use a trapezoidal rule (subroutine TRAPZD in [17, Chapter 4]) to numerically evaluate the integral arising in formula (4).

In Table 2, we summarize some of the main characteristics of A^s for the model with H_2 service times and truncation threshold $K = 150$. The existing literature shows the existence of different criteria for determining the orbit capacity K. Since the approaches in this paper are oriented to the direct computation of the mass probability function, it seems consistent to use a point criterion in order to determine the value of K. For $K = 150$, the maximum norm defined

Table 1
The value of $E[R^s]$ as a function of the service time law

ρ	μ	0.05	0.5	2.5	25.0	100.0
0.2	E_3	0.62980	0.27419	0.25466	0.25046	0.25011
	M	0.61035	0.27334	0.25450	0.25044	0.25011
	H_2	0.59722	0.27274	0.25439	0.25043	0.25010
0.4	E_3	53.9622	1.03449	0.72790	0.67255	0.66813
	M	39.6916	1.00320	0.72344	0.67213	0.66803
	H_2	33.0976	0.98513	0.72081	0.67189	0.66796
0.6	E_3	346 968.2	5.15835	1.92032	1.53751	1.50929
	M	89406.9	4.50421	1.86894	1.53335	1.50826
	H_2	44224.8	4.19806	1.84281	1.53119	1.50773
0.8	E_3	7.9274×10^{13}	85.4608	7.37906	4.25259	4.06170
	M	6.1035×10^{11}	52.5305	6.69468	4.21140	4.05183
	H_2	6.3259×10^{10}	41.8761	6.39795	4.19235	4.04724

Table 2
The main characteristics of A^s

ρ	μ	0.05	0.5	2.5	25.0	100.0
0.2	$P\{A^s = 0\}$	0.86312	0.94290	0.98422	0.99827	0.99956
	$P\{A^s \leq 100\}$	0.99998	0.99999	0.99999	0.99999	0.99999
	$E[A^s]$	1.38890	0.09098	0.01756	0.00174	0.00043
0.4	$P\{A^s = 0\}$	0.74505	0.83760	0.93888	0.99240	0.99806
	$P\{A^s \leq 100\}$	0.86801	0.99999	0.99999	0.99999	0.99999
	$E[A^s]$	48.6464	0.47770	0.08122	0.00784	0.00195
0.6	$P\{A^s = 0\}$	0.65750	0.73599	0.86983	0.98055	0.99492
	$P\{A^s \leq 100\}$	0.68505	0.99999	0.99999	0.99999	0.99999
	$E[A^s]$	29 482.2	1.79870	0.22854	0.02079	0.00515
0.8	$P\{A^s = 0\}$	0.58965	0.65154	0.78725	0.95836	0.98862
	$P\{A^s \leq 100\}$	0.60272	0.98334	0.99999	0.99999	0.99999
	$E[A^s]$	1.5814×10^{10}	9.46903	0.59948	0.04808	0.01181

by $\mathcal{P}(150) = \max_{0 \leq a \leq 100} |P\{A^s(149) = a\} - P\{A^s(150) = a\}|$ is less than 10^{-14}, where $P\{A^s(K) = a\}$ indicates that the corresponding probability is calculated from (12) after solving the system (13) and (14) with orbit capacity K. In fact, the accuracy 10^{-14} in Table 2 can be reached for lower values of K but we take $K = 150$ because this threshold guarantees that all the experiments throughout this section preserve the accuracy 10^{-14}. Firstly, we observe that the initial probability $P\{A^s = 0\}$ decreases with ρ and increases with μ. The value of $P\{A^s \leq 100\}$ shows that the tail of the distribution becomes heavier as long as ρ increases and/or μ decreases. With respect to the influence of the service time distribution, we have observed that $P\{A^s(E_3) = 0\} < P\{A^s(M) = 0\} < P\{A^s(H_2) = 0\}$ and $P\{A^s(E_3) \leq 100\} < P\{A^s(M) \leq 100\} < P\{A^s(H_2) \leq 100\}$.

For the descriptor R^s, we may comment that $P\{R^s = 0\} = \beta(\lambda)$. Moreover, our numerical experiments show that $P\{R^s(E_3) = 0\} < P\{R^s(M) = 0\} < P\{R^s(H_2) = 0\}$. As long as μ increases, the mass probability function of R^b may have two modes. One of them is always attached at the initial point $r = 0$. In contrast, the probabilities $P\{R^s = r\}$ and $P\{A^s = a\}$ are always decreasing functions of r and a, respectively.

It should be noticed that as long as ρ increases and/or μ decreases the distribution of the descriptors under study becomes more sparse. This fact is corroborated in Tables 1 and 2 for the entry $(\rho, \mu) = (0.8, 0.05)$. For this case, we observe that $P\{A^s \leq 100\} = 0.60272$ which means that the tail of the distribution has a significant weight. We also notice that the expectations take extremely large values. The dispersion of the distribution is independent of the proposed truncated approach. In fact, the dispersion at high levels of congestion is an inherent characteristic of our descriptors. It can be explained because the descriptors are referred to a busy period which becomes stochastically large as long as the congestion increases.

Table 3
Comparing the approximations of R^b

r	$\overline{F}^{E_3}_{R^b}(r)$	$\widehat{F}^{E_3}_{R^b}(r)$	$\overline{F}^{M}_{R^b}(r)$	$\widehat{F}^{M}_{R^b}(r)$	$\overline{F}^{H_2}_{R^b}(r)$	$\widehat{F}^{H_2}_{R^b}(r)$
0	0.49398	0.49205	0.55797	0.55571	0.58403	0.58160
1	0.49591	0.49207	0.56035	0.55603	0.58663	0.58199
2	0.49786	0.49210	0.56271	0.55649	0.58919	0.58257
3	0.49981	0.49216	0.56503	0.55708	0.59170	0.58330
4	0.50176	0.49226	0.56732	0.55780	0.59418	0.58417
5	0.50373	0.49241	0.56958	0.55863	0.59662	0.58518
6	0.50570	0.49261	0.57181	0.55956	0.59902	0.58631
7	0.50767	0.49288	0.57402	0.56059	0.60138	0.58755
8	0.50965	0.49323	0.57619	0.56172	0.60371	0.58889
9	0.51163	0.49366	0.57834	0.56292	0.60600	0.59032
10	0.51361	0.49418	0.58046	0.56420	0.60825	0.59183
$\leq$100	0.65630	0.66124	0.69733	0.69691	0.72417	0.72386

Table 4
Comparing the expectations of R^b

	$\overline{E}[R^b]$	$\widehat{E}[R^b]$	$E[R^b]$
E_3	1370.148958326	1370.148958321	1370.148957385
M	2042.124838966	2042.124838965	2042.124838966
H_2	2608.955483322	2608.955483324	2608.955483418

We now focus on the distribution function of R^b. The approximation based on the direct equations for the $M/PH/1/K$ retrial queue is denoted by $\overline{F}_{R^b}(r)$ (see Theorem 3), whereas $\widehat{F}_{R^b}(r)$ denotes the approach based on reallocation of the blocked primary customers (see Theorem 4).

For small values of μ, the service time tends to expire before the repeated attempt takes place. As a result, the reallocation of customers has no perturbing effect on the system dynamics. Thus, we may expect high accuracy for small retrial rates. In Table 3, we show that $\widehat{F}_{R^b}(r)$ gives also a good approximation for higher values of μ. More concretely, we choose $(\rho, \mu) = (0.8, 100.0)$ and $K = 150$. Then, the entries in the table show that $\widehat{F}_{R^b}(r)$ is close enough to $\overline{F}_{R^b}(r)$ (i.e., the exact distribution of the truncated model) for E_3, M and H_2 service times.

The above comments give some credit to our approximating assumptions (i.e., truncation of the orbit and reallocation of the blocked primary customers). In addition, the efficiency of the approximations can be measured in terms of the convergence of the approximate expected values to the true expectations given by formulas (4)–(8). In this sense, in Table 4 we supplement the results in Table 3 by comparing the corresponding expectations $\overline{E}[R^b]$ and $\widehat{E}[R^b]$, for $K = 150$, versus the true value $E[R^b]$. The equations for computing $\overline{E}[R^b]$ (respectively $\widehat{E}[R^b]$) can be easily derived by multiplying $y^b_{kj}(r)$ (respectively $P\{R^b = r\}$ and $x^b_i(r)$) by r and adding from $r = 0$ to ∞. The conclusion is that both approximations are very accurate for the three service time distributions.

Our next objective is to choose optimally the retrial rate with respect to a specified objective function. Several cost/reward criteria can be formulated. One option is to maximize the objective function:

$$f(\mu) = B_r E[R^s] + B_a E[A^s] + C_r E[R^b] + C_a E[A^b]. \quad (25)$$

The contributions of the different terms reflect the influences of the four descriptors under study. The accompanying coefficients measure the economic incidences of the profit/cost associated with each type of event. For the E_3 service times and $(B_r, B_a, C_r, C_a) = (1.0, 3.0, -10.0, -3.0)$, in Table 5 we display the value of the cost function (25). For all choices of ρ we obtain an optimal value μ^*, which has been indicated in bold in Table 5.

8. Conclusion

We have developed a computational approach for the computation of the distribution of the successful and blocked primary arrivals and repeated attempts in the $M/G/1$ retrial queue. This analysis of the customer's behavior provides

Table 5
Optimal value of μ

μ	0.05	0.5	2.5	25.0	100.0
$\rho = 0.2$	**1.4611**	−2.0860	−6.3955	−53.3011	−209.5537
$\rho = 0.4$	−257.8215	**−13.0575**	−26.2523	−192.6219	−748.1537
$\rho = 0.6$	-5.4936×10^6	−123.3615	**−111.8185**	−666.6944	−2 541.1265
$\rho = 0.8$	-3.4021×10^{15}	−4952.6061	**−922.3086**	−3722.2496	−13 709.5667

qualitative insight in the characteristics of the retrial group. The methodology can be extended to more general queuing models combining retrials and other queuing phenomena: unreliable queues [8,14,18–20], discrete-time models [8, 20], etc. The consideration of the new descriptors in a time dependent context [16] could be the subject matter of a forthcoming study. Such a study will allow us to compute our descriptors in any time interval $(0, t]$ rather than in a busy period. For a moderate time horizon t, this alternative approach should reduce the dispersion observed in Section 7 for the case $(\rho, \mu) = (0.8, 0.05)$.

Acknowledgements

The authors are grateful to the referee for his/her helpful comments on an earlier version of the paper. We acknowledge the support of DGYCIT grant MTM2005-01248.

References

[1] V.M. Abramov, Multiserver queueing systems with retrials and losses, ANZ IAM J. 48 (2007) 297–314.
[2] J. Amador, J.R. Artalejo, On the distribution of the successful and blocked events in the $M/M/c$ retrial queue: A computational approach, Appl. Math. Comput. 190 (2007) 1612–1626.
[3] J.R. Artalejo, G.I. Falin, On the characteristics of the $M/G/1$ retrial queue, Nav. Res. Logist. 47 (1996) 1147–1161.
[4] J.R. Artalejo, G.I. Falin, Standard and retrial queueing systems: A comparative analysis, Rev. Mat. Comput. 15 (2002) 101–129.
[5] J.R. Artalejo, M. Pozo, Numerical calculation of the stationary distribution of the main multiserver retrial queue, Ann. Oper. Res. 116 (2002) 41–56.
[6] J.R. Artalejo, A. Economou, M.J. Lopez-Herrero, Algorithmic analysis of the maximum queue length in a busy period for the $M/M/c$ retrial queue, INFORMS J. Comput. 19 (2007) 121–126.
[7] J.R. Artalejo, M.J. Lopez-Herrero, On the distribution of the number of retrials, Appl. Math. Mod. 31 (2007) 478–489.
[8] I. Atencia, P. Moreno, A discrete-time $Geo/G/1$ with server breakdowns, Asia–Pacific J. Oper. Res. 23 (2006) 247–271.
[9] S.R. Chakravarthy, A. Krishnamoorthy, V.C. Joshua, Analysis of multiserver retrial queue with search of customers from the orbit, Perform. Eval. 63 (2006) 776–798.
[10] J.C. Ke, H.I. Huang, C.H. Lin, On retrial queueing model with fuzzy parameters, Physica A 374 (2007) 272–280.
[11] V.I. Klimenok, A.N. Dudin, Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory, Queueing Syst. 54 (2006) 245–259.
[12] B. Krishna Kumar, J. Raja, On multiserver feedback retrial queues with balking and control retrial rate, Ann. Oper. Res. 141 (2006) 211–232.
[13] G. Latouche, V. Ramaswami, Introduction to Matrix-Analytic Methods in Stochastic Modeling, in: ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, 1999.
[14] Q.L. Li, Y. Ying, Y.Q. Zhao, A $BMAP/G/1$ retrial queue with a server subject to breakdowns and repairs, Ann. Oper. Res. 141 (2006) 233–270.
[15] E. Morozov, A multiserver retrial queue: Regenerative stability analysis, Queueing Syst. 56 (2007) 157–168.
[16] P.R. Parthasarathy, R. Sudhesh, Time-dependent analysis of a single-server retrial queue with state-dependent rates, Oper. Res. Lett. 35 (2007) 601–611.
[17] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.F. Flannery, Numerical Recipes in Fortran. The Art of Scientific Computing, Cambridge University Press, Cambridge, 1992.
[18] N.P. Sherman, J.P. Kharoufeh, An $M/M/1$ retrial queue with unreliable server, Oper. Res. Lett. 34 (2007) 697–705.
[19] J. Sztrik, B. Almasi, J. Roszik, Heterogeneous finite-source retrial queues with server subject to breakdowns and repairs, J. Math. Sci. 132 (2006) 677–685.
[20] J. Wang, Q. Zhao, Discrete-time $Geo/G/1$ retrial queue with general retrial times and starting failures, Math. Comput. Modelling 45 (2007) 853–863.
[21] X. Wu, X. Ke, Analysis of an $M/\{D_n\}/1$ retrial queue, J. Comput. Appl. Math. 200 (2007) 528–536.