On the stochastic decomposition property of single server retrial queuing systems

Nawel ARRAR[1],[*], **Natalia DJELLAB**[1], **Jean-Bernard BAILLON**[2]

[1]Probability and Statistics Laboratory, Badji Mokhtar-Annaba University, Annaba, Algeria

[2]Centre for Statistics, Analysis, and Multidisciplinary Modelling Pantheon-Sorbonne University, Paris, France

Abstract: The study of retrial queuing systems presents great analytical difficulties. Detailed results are available for some models, whereas for other models the obtained results revealed poor information and are cumbersome (they contain Laplace transforms, integral expressions, etc.). Therefore, in practice, they present limited performance. Often, to overcome this difficulty, we use an approach based on the stochastic decomposition property that can be possessed by the model. It offers the advantages of simplification of solving complex models. This paper deals with the stochastic decomposition property of an $M^X/G/1$ retrial queue with impatient customers and exponential retrial times and of an $M/G/1$ retrial system with feedback and general retrial times.

Key words: Retrial queue, server vacations, feedback, batch arrivals, approximation, stochastic decomposition property

1. Introduction

Queuing theory is a tool of stochastic modeling, performance evaluation, and supervision of real systems. It is the most suitable to provide a quantitative estimation of a system. Since the 1980s, renewed interest has been given to retrial queuing models mathematically, numerically, and in practical applications to solve the performance problems of some real systems, particularly those of telecommunication networks [15, 30, 35].

A queuing model in which an arriving customer finds all servers and all waiting positions busy and is obliged to leave the service area and try again for service after a random time is called a retrial queuing system [40]. These systems have been the subject of many studies. Recent progress is summarized in [7, 18].

There are different approaches to the study of retrial queues. We place emphasis on stochastic decomposition because it leads to simplifications when solving complex models. The general concept of the stochastic decomposition property of an $M/G/1$ queuing system is defined as follows: at a random time, the number of customers in the system is distributed as the sum of two or more independent random variables, one of which is the number of customers in the same $M/G/1$ system given that the server is idle. The mentioned systems are in the steady state. Earlier, this property was observed for queuing systems with vacations, which are characterized by the fact that the idle time of the server can be used for external tasks (priority tasks or maintenance). One can distinguish the server vacations in the case of an exhaustive service and those in the case of a nonexhaustive service. In the first situation, the server goes on vacation when the system is empty, whereas in the second situation, the server is allowed to take a vacation in the presence of customers in the system. Stochastic decomposition results for the number of customers in the system in the case of an exhaustive service

were first obtained by Fuhrmann [21] and then confirmed by Doshi [16]. A comprehensive study on vacation models in the case of a nonexhaustive service was performed by Gaver [23]. The author explicitly established the validity of the stochastic decomposition property for the models in question. However, Fuhrmann and Cooper [22] defined a series of assumptions characterizing queuing systems verifying the stochastic decomposition property, especially queuing systems with generalized vacations. When the vacation queuing model is in the steady state, the authors obtained the following decomposition result for the generating function of the steady-state distribution of the number of customers in the system at an arbitrary service completion time, $\varphi(z)$:

$$\varphi(z) = \pi(z)\chi(z),$$

where $\pi(z)$ is the generating function of the number of customers in the ordinary $M/G/1$ system without vacation, and $\chi(z)$ is the generating function of the steady-state distribution of the number of customers in the system given that the server is on vacation. This result is known as valid for any queuing system with generalized vacations and permits us to focus only on studying the effects of vacations on the number of customers in the system given that the server is on vacation.

In a sense, the retrial queuing model can be considered as a particular type of vacation model where vacation begins after each service, and its duration depends on the arrival process and system state. The stochastic decomposition for the number of customers in the $M/G/1$ retrial queue was observed by Yang and Templeton [40]. By assuming that the retrial time follows an exponential law, the authors established that the generating function of the steady-state distribution of the number of customers in the system can be written as the product of two generating functions: the first one is the Pollaczek–Khintchine equation for the number of customers in the ordinary $M/G/1$ queue, and the second one presents the generating function of the steady-state distribution of the number of customers in the $M/G/1$ retrial queue given that the server is idle. Yang et al. [39] proved that this property is always true for the general distribution of the retrial time. The considered model included the retrial mechanism depending on the number of customers in the orbit. The attempts of one orbiting customer formed a renewal process whose generic law was arbitrary. The stochastic decomposition property helps to solve some problems arising in the analysis of the $M/G/1$ retrial queue, in particular the obtaining of the factorial moments of the number of customers in the orbit and the convergence of the system in question to a limiting one when the retrial intensity goes to infinity [6].

Stochastic decomposition simplifies the resolution of the models characterized by a high interference between the components. In the case of exponential retrial times, the stochastic decomposition property was observed for retrial models with server breakdowns [2, 38]; for $M_2/G_2/1$ retrial system with two types of customers, priority and nonpriority [3, 9, 17]; for queuing systems where the phenomena of retrials and server vacations are present [5, 10]; and for retrial models with batch arrivals and persistent customers [1, 27, 40]. In the case of general retrial times and retrial mechanisms depending on the number of customers in the orbit, the validity of the property in question was verified (under some conditions) for $M/G/1$ retrial queues with server subject to active and passive breakdowns, and this was done by using an approximation method [12, 14]. In recent years, there has been an increasing interest in the study of the retrial models operating under so-called FCFS orbiting discipline (the retrial mechanism is independent of the number of customers in the orbit): orbiting customers form a FCFS queue and only the customer at the head of the queue can access the server. By assuming that the retrial time follows a general distribution, these systems were studied in the context of queuing systems with generalized vacations, that led to the establishment of their stochastic decomposition property [8, 24, 26, 29, 33, 37].

This paper deals with the stochastic decomposition property (SDP) of retrial queuing systems. To this end, we consider an $M^X/G/1$ retrial queue with impatient customers and exponential retrial times. After reviewing some results obtained in our previous work (steady-state joint distribution of the server state and the number of customers in the orbit, an embedded Markov chain) [4], we investigate the SDP of the model in question. Our study also includes the asymptotic behavior of the system under high retrial intensity. We estimate the proximity of the considered retrial queuing system and the ordinary $M^X/G/1$ queue with impatient customers. Then we demonstrate that the SDP exists in $M/G/1$ retrial queues with Bernoulli feedback and general retrial times. The obtained result is verified by using an approximation method.

The paper is organized as follows. In the next section, by assuming that the retrial time is exponentially distributed, we give the SDP of the $M^X/G/1$ retrial queue with impatient customers and study the convergence of the considered retrial model to the ordinary one (without retrials). Section 3 is devoted to the $M/G/1$ retrial queue with Bernoulli feedback. We verify its stochastic decomposition property by using an approximation method. The precision of the approximation is illustrated by some numerical results. Finally, we give the concluding remarks.

2. Stochastic decomposition for an $M^X/G/1$ retrial queue with impatient customers

In telephone networks, one can observe that a calling subscriber after some unsuccessful retrials gives up further repetitions and abandons the system (an impatient customer). In retrial queues, this phenomenon is represented by set of probabilities $\{H_k, k \geq 1\}$, where H_k is the probability that after the kth attempt fails, a customer will make the $(k+1)$th one. In general, it is assumed that the probability of a customer reinitializing after failure of a repeated attempt does not depend on the previous attempts (i.e. $H_2 = H_3 = ...$). In the queuing literature, one can find an extensive body of research addressing impatience phenomena observed in single or multiserver retrial systems but with single arrivals, e.g., [20, 28]. An $M/G/1$ retrial queue with impatient customers (where $H_2 \leq 1$ and $H_1 < 1$) was analyzed in [18]. Recent contributions on this topic include the papers of Shin and Choo [31] and Shin and Moon [32]. In [31], the authors modeled the $M/M/s$ retrial queue with balking and reneging as a Markov chain on two-dimensional lattice space $Z^+ \times Z^+$ and presented an algorithm for the steady-state distribution of the number of customers in the retrial group and service facility. The considered model contains the retrial model with finite capacity of service facility by assigning specific values to the probabilities with which the balking customers and reneging ones join the retrial group. In [32], a retrial queuing system in which the number of retrials of each customer is limited by a finite number (m) is analyzed as a model with $H_k = 1$ for $k \leq m$ and with $H_k = 0$ for $k > m$.

We consider a single server queuing system at which primary customers arrive in batches of size k (with probability c_k, $k \geq 1$) according to a Poisson stream with rate $\lambda > 0$. If the server is busy at the arrival epoch, then with probability $1 - H_1 > 0$ all these customers leave the system without receiving service and with probability H_1 join the retrial group (orbit), whereas if the server is idle, then one of the arriving customers begins his service and leaves the system after the service, while the rest of the batch's customers go to the orbit. Let $C(z) = \sum_{k=1}^{\infty} c_k z^k$ be the generating function of the steady-state distribution of the batch size and $\bar{c} = C'(1)$ be the mean batch size. The impatience phenomenon is represented by the probability $1 - H_1$ (the probability $H_2 = 1$). Any orbiting customer will repeatedly retry until the time at which he finds the server idle and starts his service. The retrial times are exponentially distributed with distribution function $T(x) = 1 - e^{-\theta x}$, $x \geq 0$, having finite mean $\frac{1}{\theta}$. The service times follow a general distribution with distribution

function $B(x)$ and Laplace–Stieltjes transform $\widetilde{B}(s)$, $Re(s) > 0$. Let $\beta_k = (-1)^k \widetilde{B}^{(k)}(0)$ be the kth moment of the service time about the origin and $\rho = \lambda \bar{c} H_1 \beta_1$ be the traffic intensity. Finally, we admit the hypothesis that all random variables defined above are mutually independent.

This model can be used to study a situation that is frequently observed in communication networks and characterized by the fact that a batch of packets that is not taken for service immediately upon arrival can leave the system once and for all or join the "orbit" from which it retries to be served at random time intervals.

2.1. Steady-state distribution of the system state

The state of the considered system at time t can be described by means of the process $\{C(t), N_0(t), \zeta(t), t \geq 0\}$, where $N_0(t)$ is the number of customers in the orbit and $C(t)$ is the state of the server at time t. We have that $C(t)$ is 0 or 1 depending on whether the server is idle or busy. If $C(t) = 1$, $\zeta(t)$ represents the elapsed service time of the customer in service at time t.

Let $\rho < 1$. Define

$$P_{0,n} = \lim_{t \to \infty} P(C(t) = 0, N_0(t) = n);$$

$$P_{1,n} = \int_0^\infty \lim_{t \to \infty} \frac{d}{dx} P(C(t) = 1, \zeta(t) \leq x, N_0(t) = n) dx.$$

In our previous work [4], we found the partial generating functions:

$$P_0(z) = \sum_{n=0}^\infty z^n P_{0,n} = \frac{H_1(1-\rho)}{\rho + H_1(1-\rho)} \exp\left[\frac{\lambda}{\theta} \int_1^z \frac{1 - g(u)\frac{C(u)}{u}}{g(u) - u} du\right]$$

and

$$P_1(z) = \sum_{n=0}^\infty z^n P_{1,n} = \int_0^\infty P_1(z,x) dx = \frac{1 - g(z)}{g(z) - zH_1} P_0(z),$$

where $g(u) = \widetilde{B}(\lambda H_1(1 - C(u)))$; the generating function of the number of customers in the orbit is

$$P(z) = P_0(z) + P_1(z),$$

$$= \frac{(1 - H_1 z) - g(z)(1 - H_1)}{H_1(g(z) - z)}$$

$$\times \frac{H_1(1-\rho)}{\rho + H_1(1-\rho)} \exp\left[\frac{\lambda}{\theta} \int_1^z \frac{1 - g(u)\frac{C(u)}{u}}{g(u) - u} du\right]$$

and the steady-state distribution of the server state is

$$P_0 = \lim_{t \to \infty} P(C(t) = 0) = P_0(1) = \frac{H_1(1-\rho)}{\rho + H_1(1-\rho)},$$

$$P_1 = \lim_{t \to \infty} P(C(t) = 1) = P_1(1) = \frac{\rho}{\rho + H_1(1-\rho)}.$$

921

We have also studied the stationary distribution of the number of customers in the orbit at departure times ξ_k (the time when the server becomes idle for the kth time). The generating function of the steady-state distribution $\pi_n = \lim_{k \longrightarrow \infty} P\left(N_0\left(\xi_k\right) = n\right)$ of the embedded Markov chain was obtained as follows [4]:

$$\varphi\left(z\right) = \frac{1-\rho}{\overline{c}\rho} \frac{g\left(z\right)\left(1 - C\left(z\right)\right)}{g\left(z\right) - z} \exp\left[\frac{\lambda}{\theta} \int_1^z \frac{1 - g\left(u\right)\frac{C\left(u\right)}{u}}{g\left(u\right) - u} du\right]. \tag{1}$$

2.2. Stochastic decomposition property

At present, we investigate the decomposition property of the considered system. In the context of our paper, vacations of the server are due to the retrials and the model without vacations is the queuing system with batch arrivals, waiting line, and impatience phenomenon.

From (1), $\varphi\left(z\right)$ can be rewritten as a product of two factors:

$$\varphi\left(z\right) = \Omega\left(z\right) \Psi\left(z\right). \tag{2}$$

We have found that the first factor $\Omega\left(z\right) = \frac{1-\rho}{\overline{c}\rho} \frac{g(z)(1-C(z))}{g(z)-z}$ is the generating function for the number of customers at the departure epochs in the ordinary $M^X/G/1$ queue with impatient batches (model M_∞). The second factor $\Psi\left(z\right) = \frac{P_0(z)}{P_0(1)} = \exp\left\{\frac{\lambda}{\theta} \int_1^z \frac{1-g(u)\frac{C(u)}{u}}{g(u)-u} du\right\}$ presents the generating function for the number of customers at departure epochs in the corresponding retrial queue (model M_θ) given that the server is idle. Thus, the SDP of the considered system can be expressed in the following manner:

$$\{0, N_{o,\theta}(t), t \geq 0\} = \{0, N_{q,\infty}(t), t \geq 0\} + \{0, R_\theta(t), t \geq 0\}. \tag{3}$$

The processes $\{0, N_{o,\theta}(t), t \geq 0\}$ and $\{0, R_\theta(t), t \geq 0\}$ are related to the model M_θ, where $R_\theta(t)$ represents the number of customers in the orbit at time t given that the server is idle. The process $\{0, N_{q,\infty}(t), t \geq 0\}$ is associated with the model M_∞, where $N_{q,\infty}(t)$ is the number of customers in the waiting line at time t.

Therefore, the number of customers in the $M^X/G/1$ retrial queue with impatient customers (at the time when the server enters the idle state) is equal to the sum of two independent random variables: the number of customers in the ordinary $M^X/G/1$ queue with impatient batches (at idle epochs of the server) and the number of customers in the corresponding retrial queue given that the server is idle. The obtained result is important for understanding the real contribution of the retrials to the number of customers in the system.

2.3. Behavior of the system under high retrial rate

Now we study the asymptotic behavior of our system when the retrial rate is high. When $\theta \to \infty$, the steady-state distribution of the system with retrials converges to a limit, which is generally the steady-state distribution of a limiting system. In our case, it is intuitive that this is the model M_∞. To prove this heuristic argument, we use the SDP.

Let

$$
\begin{aligned}
\pi_n(\theta) &= \lim_{k\to\infty} P\left(N_{o,\theta}(\xi_k) = n\right), \\
\pi_n(\infty) &= \lim_{k\to\infty} P\left(N_{q,\infty}(\xi_k) = n\right), \\
q_n\left(\theta\right) &= \lim_{k\to\infty} P(N_\theta(\xi_k) = n).
\end{aligned} \tag{4}
$$

Note that $q_n(\theta) = \frac{P_{0,n}}{\left(\frac{H_1(1-\rho)}{\rho + H_1(1-\rho)}\right)}$ and its generating function is $\Psi(z)$. We can announce the following result:

Theorem 1 *As $\theta \to \infty$ the distance*

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)|$$

between distributions $\pi_n(\theta)$ and $\pi_n(\infty)$ is $o\left(\frac{1}{\theta}\right)$. To be more exact, the following inequalities hold:

$$2\frac{1-\rho}{\overline{c}\rho \widetilde{B}(\lambda H_1)}(1 - q_0(\theta)) \le \sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \le 2(1 - q_0(\theta)),$$

where

$$q_0(\theta) = \exp\left\{-\frac{\lambda}{\theta} \int_0^1 \frac{1 - g(u)\frac{C(u)}{u}}{g(u) - u} du\right\}.$$

Proof From (2), we see that $\pi_n(\theta)$ is a convolution of two distributions, $\pi_n(\infty)$ and $q_n(\theta)$; that is,

$$\pi_n(\theta) = \sum_{k=0}^{n-1} \pi_k(\infty) q_{n-k}(\theta). \tag{5}$$

Expression (5) can be rewritten as

$$\pi_n(\theta) - \pi_n(\infty) = \pi_n(\infty)q_0(\theta) - \pi_n(\infty) + \sum_{k=0}^{n} \pi_k(\infty)q_{n-k}(\theta).$$

We obtain that

$$
\begin{aligned}
|\pi_n(\theta) - \pi_n(\infty)| &\le |\pi_n(\infty)q_0(\theta) - \pi_n(\infty)| + \sum_{k=0}^{n-1} \pi_k(\infty)q_{n-k}(\theta) \\
&\le \pi_n(\infty)(1 - 2q_0(\theta)) + \pi_n(\theta),
\end{aligned}
$$

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \le 2(1 - q_0(\theta)),$$

where

$$q_0(\theta) = \Psi(0) = \exp\left\{-\frac{\lambda}{\theta} \int_0^1 \frac{1 - g(u)\frac{C(u)}{u}}{g(u) - u} du\right\}.$$

Thus, the upper inequality follows. By using the inequality $|x - y| \ge x - y$, we find that

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \ge |\pi_0(\theta) - \pi_0(\infty)| + \pi_0(\infty) - \pi_0(\theta). \tag{6}$$

From (5), one can see that $\pi_0(\theta) = \pi_0(\infty)q_0(\theta) < \pi_0(\infty)$. Therefore, expression (6) becomes

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \geq 2\pi_0(\infty)(1 - q_0(\theta)).$$

We obtain the probability $\pi_0(\infty)$ from the generating function $\frac{1-\rho}{\bar{c}\rho} \frac{(1-C(z))}{g(z)-z}$ of the random variable $N_{q,\infty}$ by putting $z = 0$; that is, $\pi_0(\infty) = \frac{1-\rho}{\bar{c}\rho\widetilde{B}(\lambda H_1)}$.

Thus,

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \geq 2\frac{1-\rho}{\bar{c}\rho\widetilde{B}(\lambda H_1)}(1 - q_0(\theta)).$$

<div align="right">□</div>

2.4. Numerical illustration

In Table 1 below, we present some numerical values of the lower bounds and upper bounds of the inequalities verified by the number of customers in the system when the retrial rate θ tends to ∞. We assign the following values to the various parameters of the $M^X/G/1$ retrial queue with batch arrivals and impatient customers: $\lambda = 0.2, \gamma = \frac{1}{\beta_1}, H_1 = 0.8$ (such that $\rho = 0.8$). Suppose that the service time follows the law:

1. Exponential E: $B(x) = 1 - e^{-\gamma x}, x \geq 0$, with coefficient of variation $CV = 1$;

2. Two-stage Erlang E_2: $B(x) = 1 - e^{-2\gamma x} - 2\gamma x e^{-2\gamma x}, x \geq 0$, with coefficient of variation $CV \simeq 0.7$;

3. Two hyperexponential H_2: $B(x) = 1 - p_1 e^{-\gamma_1 x} - p_2 e^{-\gamma_2 x}, x \geq 0$, where $p_1 + p_2 = 1$ and $\frac{p_1}{\gamma_1} + \frac{p_2}{\gamma_2} = \frac{1}{\gamma}$. The coefficient of variation $CV = 1.5$, and then $p_1 \simeq 0.19$, $p_2 \simeq 0.81$, and $\gamma_1 \simeq 0.38, \gamma_2 \simeq 2$.

We can see that the convergence is faster when the service time is exponential.

Table 1. $M^X/G/1$ retrial queue with impatient customers.

θ	Exponential		Two-stage Erlang		Two-stage hyperexponential	
	Lower bound (LB)	Upper bound (UB)	Lower bound (LB)	Upper bound (UB)	Lower bound (LB)	Upper bound (UB)
1	0.0029800	0.05138	0.0558671	0.95794	0.1943957	0.93146
5	0.0006020	0.01038	0.0142592	0.24450	0.0491822	0.23566
10	0.0003016	0.00520	0.0073623	0.12624	0.0253612	0.12152
100	0.0000302	0.00052	0.0007582	0.01300	0.0026088	0.01250
10^3	0.0000035	0.00006	0.0000758	0.00130	0.0002630	0.00126
10^4	0	0	0.0000082	0.00014	0.0000250	0.00012
10^5	0	0	0.0000012	0.00002	0.0000042	0.00002
10^6	0	0	0	0	0	0

3. Stochastic decomposition for $M/G/1$ retrial queue with feedback and general retrial time

Most of the papers on retrial queues have considered queuing systems without feedback, but many practical situations (for instance, in communication networks where data transmissions need to be guaranteed to be error-free within a specified probability, or feedback schemes that are used to request retransmission of packets that are lost or received in a corrupted form) can be modeled as retrial queues with feedback. The analysis of these systems can be found in [11, 19, 25, 26].

We consider a single server queuing system with no waiting space. The primary customers arrive according to a Poisson process with rate $\lambda > 0$. An arriving customer receives immediate service if the server is idle; otherwise, he leaves the service area temporarily to join the retrial group (orbit). Times between the successive retrials of any orbiting customer are governed by an arbitrary probability distribution function $T(x)$ having finite mean $\frac{1}{\theta}$. The service times follow a general distribution with distribution function $B(x)$ having finite mean $\frac{1}{\gamma}$ and Laplace–Stieltjes transform $\widetilde{B}(x)$. After the customer is served, he will decide either to join the orbit for another service with probability c or to leave the system forever with probability $\bar{c} = 1 - c$. Finally, we admit the hypothesis of mutual independence between all random variables defined above.

The state of the system at time t can be described by means of the process $\{C(t), N_o(t), \zeta(t), \in(t), t \geq 0\}$, where $N_o(t)$ is the number of customers in the orbit and $C(t)$ is the state of the server at time t. We have that $C(t)$ is 0 or 1 depending on whether the server is idle or busy. If $C(t) = 1$, $\zeta(t)$ represents the elapsed service time of the customer being served. When $C(t) = 0$ and $N_o(t) > 0$, the random variable $\zeta(t)$ represents the elapsed retrial time.

3.1. Notations

Let ξ_n be the time when the server enters the idle state for the nth time, ς be the time at which the nth customer arrives at the server, X_i^n be the time elapsed since the last attempt made by the ith customer in the orbit until instant ξ_n^+, $and\, q_n = N_o(\xi_n^+)$ be the number of customers in the orbit at instant ξ_n^+.

We assume that the system is in steady state; that is, $\rho = \frac{\lambda}{\gamma} + c < 1$ [15]. Let $q = \lim_{n \to \infty} q_n; X_i = \lim_{n \to \infty} X_i^n$. When $q > 0$, we have a vector $X = (X_1, X_2, ..., X_q)$. We denote by $f_q(x_1, x_2, ..., x_q) = f_q(x)$ the joint density function of q and X.

Define

$$r_{ij} = \lim_{n \to \infty} P(C(\varsigma_n^-) = i, N_o(\varsigma_n^-) = j) \quad i = 0, 1 \; j = 0, 1, 2, ...$$

$$p_{ij} = \lim_{n \to \infty} P(C(t) = i, N_o(t) = j) \quad i = 0, 1 \; j = 0, 1, 2, ...$$

$$d_k = \lim_{n \to \infty} P(N_o(\xi_n^+) = k) \quad k = 0, 1, 2, ...; d_k = \int_0^\infty f_k(x)\, dx \quad k = 1, 2, ...$$

$$D(z) = \sum_{k=0}^\infty d_k z^k \text{ and } R_i(z) = \sum_{j=0}^\infty r_{ij} z^j \; i = 0, 1, ...$$

3.2. Stochastic decomposition property

The decomposition property of the $M/G/1$ retrial queue with Bernoulli feedback and exponential retrial times (linear retrial policy) was analyzed in [13]. Now we extend the obtained result for general retrial time distribution. To this end, the method of the embedded Markov chain is used.

Consider a sequence of random variables $\{q_n = N_o(\xi_n^+), n \geq 1\}$. This is an embedded Markov chain for our model. Its fundamental equation is

$$q_{n+1} = q_n - \delta(q_n; X^n) + v_{n+1} + u, \tag{7}$$

where $\delta(q_n; X^n)$ is 1 or 0 depending on whether the $(n+1)$th served customer is an orbiting customer or a primary one. When $q_n = 0, P(\delta(0; X^n) = 0) = P(\delta(0) = 0) = 1$. The random variable v_{n+1} represents the number of primary customers arriving at the system during the $(n+1)$th service time interval. It does not depend on events that have occurred before the beginning of the $(n+1)$th service. Its distribution is given by

$$k_i = P(v_{n+1} = i) = \int_0^\infty \frac{(\lambda x)^i}{i!} e^{-\lambda x} dB(x),$$

having the generating function $K(z) = \sum_{i=0}^\infty k_i z^i = \widetilde{B}(\lambda - \lambda z)$ [6]. The random variable u is 0 or 1 depending on whether the served customer leaves the system or goes to orbit. We have also that $P(u = 0) = \bar{c}$ and $P(u = 1) = c$.

Theorem 2 *If the $M/G/1$ retrial queue with Bernoulli feedback and general retrial times is in the steady state $\left(\rho = \frac{\lambda}{\gamma} + c < 1\right)$, we have the following decomposition result for the generating function of the steady-state distribution of the embedded Markov chain, $D(z)$:*

$$D(z) = \frac{(1 - \rho)\widetilde{B}(\lambda - \lambda z)(1 - z)}{(\bar{c} + cz)\widetilde{B}(\lambda - \lambda z) - z} \cdot \frac{(\bar{c} + cz)R_0(z)}{1 - \rho}.$$

Proof Consider the fundamental equation (7). Since the random variables $v_{n+1}, q_n - \delta(q_n; X^n)$, and u are mutually independent, we have

$$z^{q_{n+1}} = z^{q_n - \delta(q_n; X^n)} z^{v_{n+1}} z^u;$$

$$\mathbb{E}\left[z^{q_{n+1}}\right] = \mathbb{E}\left[z^{q_n - \delta(q_n; X^n)}\right] \mathbb{E}\left[z^{v_{n+1}}\right] \mathbb{E}\left[z^u\right].$$

Let $n \to \infty$. We find

$$D(z) = \mathbb{E}\left[z^{q - \delta(q; X)}\right] \widetilde{B}(\lambda - \lambda z)(\bar{c} + cz). \tag{8}$$

Using the rule of conditional expectation, one can obtain

$$
\begin{aligned}
\mathbb{E}\left[z^{q-\delta(q;X)}\right] &= \sum_{j=0}^{\infty}\int_0^{\infty}f_j\left(x\right)\mathbb{E}\left[z^{j-\delta(j;x)}\right]dx \\
&= \sum_{j=0}^{\infty}\int_0^{\infty}f_j\left(x\right)\left[z^j P\left(\delta\left(j;x\right)=0\right)+z^{j-1}\left(1-P\left(\delta\left(j;x\right)=0\right)\right)\right]dx \\
&= \sum_{j=0}^{\infty}\left[\int_0^{\infty}f_j\left(x\right)z^j P\left(\delta\left(j;x\right)=0\right)dx+\frac{1}{z}\int_0^{\infty}f_j\left(x\right)z^j dx\right. \\
&\qquad\left.-\frac{1}{z}\int_0^{\infty}f_j\left(x\right)z^j P\left(\delta\left(j;x\right)=0\right)dx\right] \\
&= \frac{1}{z}\sum_{j=0}^{\infty}z^j\int_0^{\infty}f_j\left(x\right)dx+\left(1-\frac{1}{z}\right)\sum_{j=0}^{\infty}z^j\int_0^{\infty}f_j\left(x\right)P\left(\delta\left(j;x\right)=0\right)dx \\
&= \frac{1}{z}\sum_{j=0}^{\infty}z^j d_j+\left(1-\frac{1}{z}\right)\sum_{j=0}^{\infty}z^j\int_0^{\infty}f_j\left(x\right)P\left(\delta\left(j;x\right)=0\right)dx.
\end{aligned}
\tag{9}
$$

Consider $\int_0^{\infty}f_j\left(x\right)P\left(\delta\left(j;x\right)=0\right)dx$. This is the probability that an arriving customer finds j customers in the orbit and no customer at the server. This event takes place if and only if the last served customer leaves j customers in the orbit, he still did not decide to join the orbit or to leave the system, and the new arrival occurs before any of the j orbiting customers retry for service. Therefore, $r_{oj}=\int_0^{\infty}f_j\left(x\right)P\left(\delta\left(j;x\right)=0\right)dx$. We can rewrite (9) as

$$
\mathbb{E}\left[z^{q-\delta(q;X)}\right]=\frac{1}{z}D\left(z\right)+\left(1-\frac{1}{z}\right)R_0\left(z\right).
\tag{10}
$$

Finally, putting (10) into (8), we obtain

$$
D\left(z\right)=\left[\frac{1}{z}D\left(z\right)+\left(1-\frac{1}{z}\right)R_0\left(z\right)\right]\widetilde{B}\left(\lambda-\lambda z\right)\left(\bar{c}+cz\right)
$$

or

$$
D\left(z\right)=\frac{\left(1-\rho\right)\widetilde{B}\left(\lambda-\lambda z\right)\left(1-z\right)}{\left(\bar{c}+cz\right)\widetilde{B}\left(\lambda-\lambda z\right)-z}\cdot\frac{\left(\bar{c}+cz\right)R_0\left(z\right)}{1-\rho}.
\tag{11}
$$

$\square$

One can see that the first factor on the right-hand part of (11) is the generating function for the number of customers in the $M/G/1$ queuing system with Bernoulli feedback [34]; the remaining one is the generating function for the number of customers in the retrial queue with feedback given that the server is idle. Note that if $\bar{c}=1$, one can obtain the same result as in [39] for the $M/G/1$ retrial queue without feedback.

3.3. Approximate solution

From (11), it is easy to see that steady-state distribution $\{d_k,k\geq 0\}$ of the embedded Markov chain is a convolution of two distributions: the steady-state queue size distribution for a model without retrials (we denote by $\{a_k,k\geq 0\}$) and the steady-state joint distribution $\{p_{0,k},k\geq 0\}$.

Obviously $d_k = P(q = k) = \int_0^\infty f_k(x)\,dx$ for $k > 0$. Since Poisson arrivals see time averages [36], we have $p_{0,k} = r_{0,k}^*$ for $k \geq 0$, where $r_{0,k}^* = \bar{c}r_{0,k} + cr_{0,k-1}$ is defined by the generating function $(\bar{c} + cz)\,R_0(z)$.

Since expanded retrial times $X_1, X_2, \cdots, X_k$ of the $k > 0$ orbiting customers depend on each other in a very complicated way, a derivation of an explicit formula for the joint density function $f_k(x)$ is difficult, if not impossible.

An approximation to $f_k(x)$ was proposed in [39]: $f_k(x) \approx d_k \theta^k \prod_{i=1}^k (1 - T(x_i))$. It is based on the intuitive consideration that mean retrial time is very small relative to the mean service time. Using the above approximation, it was established that $r_{0,k}^* \approx d_k b_k$, where $b_k = \int_0^\infty (1 - m(t))^k \lambda e^{-\lambda t} dt$ with $m(t) = \int_0^t \theta (1 - T(u))\,du$.

We assume that $\{a_k, k \geq 0\}$ is already known. Under this assumption, we can express the result (11) in the following manner:

$$d_k = \frac{1}{1 - \rho} \sum_{i=0}^k a_i r_{0,k-i}^* \tag{12}$$

with

$$r_{0,k}^* \approx d_k b_k, \tag{13}$$

$$\sum_{k=0}^\infty d_k = 1. \tag{14}$$

The set of equations (12)–(14) gives an approximate solution to $\{d_k, k \geq 0\}$, respectively. From (12)–(14) and using $a_0 = 1 - \rho$, it is easy to find the following computational procedure:

$$\widehat{d}_k = g_k \widehat{d}_0 \qquad k = 0, 1, \cdots; \tag{15}$$

$$\widehat{d}_0 = \frac{1}{\sum_{k=0}^\infty g_k};$$

where

$$g_0 = 1; \quad g_k = \frac{1}{(1-\rho)(1-b_k)} \sum_{i=1}^k a_i b_{k-i} g_{k-i} \qquad k = 1, 2, \cdots.$$

Once the steady-state probabilities $\left\{\widehat{d}_k, k \geq 0\right\}$ are evaluated, we can calculate the mean number of customers in the system at an arbitrary time when the server is able to start a new service time, $\mathbb{E}[N] : \mathbb{E}[N] \approx \sum_{k=0}^\infty k\widehat{d}_k$, and also the variance $\mathbb{V}[N] : \mathbb{V}[N] \approx \sum_{k=0}^\infty k^2\widehat{d}_k - (\mathbb{E}[N])^2$.

In [39], for $M/G/1$ retrial queues (without feedback), it was shown that the performance of the approximation is not affected very much by the type of service time distribution (or its coefficient of variation cs). The approximation performs well as long as the mean retrial time is less than the mean service time and the coefficient of variation of retrial times cv is fairly close ($cv \leq 4$) to that of the exponential distribution (for which the approximation produces the exact solution). The mean number of customers in the system at an arbitrary time when the server is able to start a new service time $\mathbb{E}[N]$ is an increasing function of the second moments of both the service time distribution and the retrial time distribution.

Table 2. M/M/1 retrial queue with feedback: $\gamma = 1, c = 0.1$.

			Retrial times		
			E	E_2	H_2
λ	θ		$\mathbb{E}[N]$		
0.1	1	Approx	0.1510	0.1478	0.1580
		Simul	0.1494	0.1437	0.1577
	3.3	Approx	0.1414	0.1403	0.1437
		Simul	0.1394	0.1388	0.1398
	10	Approx	0.1385	0.1382	0.1393
		Simul	0.1334	0.1306	0.1371
λ	θ		$\mathbb{E}[N]$		
0.3	1	Approx	0.7023	0.6682	0.7562
		Simul	0.7093	0.6759	0.7716
	3.3	Approx	0.5862	0.5752	0.6080
		Simul	0.5654	0.5603	0.5815
	10	Approx	0.5523	0.5486	0.5603
		Simul	0.5400	0.5361	0.5426
λ	θ		$\mathbb{E}[N]$		
0.6	1	Approx	3.4022	3.2163	3.5135
		Simul	3.4750	3.2192	4.1864
	3.3	Approx	2.4729	2.4061	2.5565
		Simul	2.4450	2.3416	2.6404
	10	Approx	2.2022	2.1790	2.2384
		Simul	2.1692	2.1151	2.2395

Table 3. $M/M/1$ retrial models: $\gamma = 1$.

λ	θ	Retrial times							
		E	E	E_2	E_2	E_2	H_2	H_2	H_2
		M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3
		$\mathbb{E}[N]$	$\mathbb{E}[N]$	$\mathbb{E}[N]$	$\mathbb{E}[N]$	$\mathbb{E}[N]$	$\mathbb{E}[N]$	$\mathbb{E}[N]$	$\mathbb{E}[N]$
		$\mathbb{V}[N]$	$\mathbb{V}[N]$	$\mathbb{V}[N]$	$\mathbb{V}[N]$	$\mathbb{V}[N]$	$\mathbb{V}[N]$	$\mathbb{V}[N]$	$\mathbb{V}[N]$
0.1	1	0.1262	0.1510	0.1194	0.1234	0.1478	0.1277	0.1320	0.1580
		0.1399	0.1682	0.1323	0.1370	0.1647	0.1410	0.1460	0.1756
	3.3	0.1181	0.1414	0.1135	0.1173	0.1403	0.1164	0.1201	0.1437
		0.1310	0.1574	0.1258	0.1301	0.1562	0.1293	0.1331	0.1598
	10	0.1158	0.1385	0.1119	0.1155	0.1382	0.1128	0.1165	0.1393
		0.1284	0.1542	0.1242	0.1281	0.1338	0.1251	0.1291	0.1550
0.3	1	0.5684	0.7023	0.5304	0.5410	0.6682	0.5997	0.6126	0.7562
		0.8102	1.0344	0.7597	0.7749	0.9891	0.8411	0.8610	1.0969
	3.3	0.4765	0.5862	0.4591	0.4677	0.5752	0.4842	0.4943	0.6080
		0.6788	0.8601	0.6585	0.6673	0.8453	0.6864	0.6997	0.8863
	10	0.4498	0.5523	0.4385	0.4468	0.5486	0.4484	0.4562	0.5603
		0.6405	0.8094	0.6268	0.6366	0.8044	0.6515	0.6482	0.8190
0.6	1	2.4414	3.4022	2.2623	2.3006	3.2163	2.5471	2.5707	3.5135
		6.1069	10.0639	5.7474	5.8466	9.6750	6.1704	6.0895	9.5331
	3.3	1.7988	2.4729	1.7204	1.7492	2.4061	1.8413	1.8710	2.5565
		4.4950	7.2759	4.3088	4.3981	7.1271	4.5037	4.5704	7.2894
	10	1.6117	2.2022	1.5732	1.5945	2.1790	1.6167	1.6414	2.2384
		4.0255	6.4639	3.9419	3.9914	6.4110	3.9748	4.0631	6.4931

We have examined the performance of the approximation in the case of $M/G/1$ (M/M/1, M/E$_2$/1, and M/H$_2$/1) retrial queues with feedback and obtained the same conclusions as in [39]. For illustration purposes, we present the numerical results only for the M/M/1 model.

For retrial time distributions, we have exponential (E) with $cv = 1$, two-stage Erlang (E$_2$) with $cv \approx 0.7$, and two-stage hyperexponential (H$_2$) with $cv = 1.5$. Table 2 contains the approximate results calculated according to (15) and those from a simulation study (at 95% confidence intervals). We observe that the approximate results are close to the simulation ones even when the mean retrial times are as large as the mean service time.

At present, we compare the efficacy of the approximation used for different models: M$_1$− without feedback [39], M$_2$−with breakdowns [12], and M$_3$−with feedback.

From Table 3 we can observe that the approximate results obtained for these models present the same tendency not only for the mean number of customers in the system at an arbitrary time when the server is able to start a new service time $E[N]$ but also for the variance $V[N]$. Furthermore, $E[N]$ is an increasing function of the second moment of retrial time distribution.

4. Concluding remarks

In this paper, we have established the stochastic decomposition property of two single retrial queues: an $M^X/G/1$ queuing system with impatient batches and exponential retrial times and also an $M/G/1$ queue with feedback and general retrial times. In the first case, the property in question was used to study the convergence of the considered model with retrials to the ordinary one (without retrials). In the second case, by using the established property, we have presented an approximation method that permits us to conclude that the number of customers in the $M/G/1$ retrial queue with feedback (at idle epochs of the server) is equal to the sum of two independent random variables: the number of customers in the ordinary $M/G/1$ queue with feedback and the number of customers in the retrial queue with feedback given that the server is idle.

References

[1] Aissani A. An M^X/G/1 energetic retrial queue with vacations and its control. Electron Notes Theor Comput Sci 2009; 253: 33-44.

[2] Aissani A, Artalejo JR. On the single server retrial queue subject to breakdowns. Queueing Syst 1998; 30: 309-321.

[3] Arivudainambi D, Averbakh I, Berman O. Stationary analysis of a single server retrial queue with priority and vacation. Int J Oper Res 2009; 5: 26-47.

[4] Arrar NK, Djellab N, Baillon JB. On the asymptotic behavior of $M/G/1$ retrial queues with batch arrivals and impatience phenomenon. Math Comput Model 2012; 55: 654-665.

[5] Artalejo JR. Analysis of an $M/G/1$ queue with constant repeated attempts and server vacations. Comput Oper Res 1997; 24: 493-504.

[6] Artalejo JR, Falin GI. Stochastic decomposition for retrial queues. TOP 1994; 2: 329-342.

[7] Artalejo JR, Gomez-Corral A. Retrial Queueing Systems: A Computational Approach. Berling, Germany: Springer 2008.

[8] Atencia I, Fortes I, Sánchez S. A discrete-time retrial queueing system with starting failures, Bernoulli feedback and general retrial times. Comput Ind Eng 2009; 57: 1291-1299.

[9] Atencia I, Moreno P. A single server retrial queue with general retrial times and Bernoulli schedule. Appl Math Comput 2005; 162: 855-880.

[10] Boualem M, Djellab N, Aissani D. An $M/G/1$ retrial queue with exhaustive service and server vacations. J Com Comput 2011; 8: 720-726.

[11] Choi BD, Kim YC, Lee YW. The M/M/c retrial queue with geometric loss and feedback. Comput Math Appl 1998; 36: 41-52.

[12] Djellab NV. On the $M/G/1$ retrial queue subjected to breakdowns. RAIRO-Oper Res 2002; 36: 299-310.

[13] Djellab NV. On the $M/G/1$ retrial queue with feedback. In: Proceedings of the International Conference 11Mathematical Methods of Optimization of Telecommunication Networks", Minsk 2005, pp. 32-35.

[14] Djellab NV. On the single server retrial queue. Yugoslav Journal of Operations Research 2006; 16: 45-53.

[15] Do TV. Solution for retrial queueing problem in cellular networks with fractional guard channel policy. Math Comput Model 2011; 53: 2059-2066.

[16] Doshi BT. Queueing systems with vacations: a survey. Queueing Syst 1986; 1: 29-66.

[17] Falin GI, Artalejo JR, Martin M. On the single server retrial queue with priority customers. Queueing Syst 1993; 14: 439-455.

[18] Falin GI, Templeton JGC. Retrial Queues. New York, NY, USA: Chapman and Hall 1997.

[19] Farahmand K. Single line queue with recurrent repeat demands. Queueing Syst 1996; 22: 425-435.

[20] Fayolle G, Brun MA. On a system with impatience and repeated calls. In: Boxma OJ, editor. Queueing Theory and its Applications, Liber Amicorum for J.W. Cohen. Amsterdam, the Netherlands: North Holland, 1988, pp. 283-305.

[21] Fuhrmann SW. Performance Analysis of a Class of Cyclic Schedules. Bell Laboratories Technical Memorandum 59531-1. Murray Hill, NJ, USA: Bell Laboratories, 1981.

[22] Fuhrmann SW, Cooper RB. Stochastic decompositions in the $M/G/1$ queue with generalized vacations. Oper Res 1985; 33: 1117-1129.

[23] Gaver DP. A waiting line with interrupted service including priorities. J Roy Stat Soc B 1962; 25: 73-90.

[24] Gomez-Corral A. Stochastic analysis of a single server retrial queue with general retrial times. Nav Res Log 1999; 46: 561-581.

[25] Krishna Kumar B, Arivudainambi D. On the busy period of an $M/G/1$ Bernoulli feedback retrial queue with negative customers and preemptive resume. In: Artalejo JR, Krishnamoorthy A, editors. Advances in Stochastic Modeling. Branchburg, NJ, USA: Notable Publications, 2002, pp. 205-218.

[26] Krishna Kumar B, Pavai Madheswari S, Vijayakumar. The $M/G/1$ retrial queue with feedback and starting failure. Appl Math Model 2002; 26: 1057-1075.

[27] Langaris C, Moutzoukis E. A retrial queue with structured batch arrivals, priorities and server vacations. Queueing Syst 1995; 20: 341-368.

[28] Martin M, Artalejo JR. Analysis of an $M/G/1$ queue with two types of impatient units. Adv Appl Probab 1995; 27: 840-861.

[29] Moreno P. An $M/G/1$ retrial queue with recurrent customers and general retrial times. Appl Math Comput 2004; 159: 651-666.

[30] Phung-Duc T, Kawanishi K. Performance analysis of call centers with abandonment, retrial and after-call work. Perform Evaluation 2014; 80: 43-62.

[31] Shin YW, Choo TS. $M/M/s$ queue with impatient customers and retrials. Appl Math Model 2009; 33: 2596-2606.

[32] Shin YW, Moon DH. Retrial queues with limited number of retrials: numerical investigations. In: 7th International Symposium on Operations Research and Its Applications, Lijiang 2008, pp. 237-247.

[33] Sumitha D, Chandrika KU. $M^X/G/1$ retrial queue with second optional service, feedback, admission control and vacation. Int J Innov Res Sci Eng Tech 2014; 3: 8320-8329.

[34] Takacs I. A single server queue with feedback. Bell Syst Tech J 1963; 42: 505-519.

[35] Tran-Gia P, Mandjes M. Modeling of customer retrial phenomenon in cellular mobile networks. IEEE J Sel Area Comm 1997; 15: 1406-1411.

[36] Wolff R. Poisson arrivals see time averages. Oper Res 1982; 30: 223-231.

[37] Wu J, Wang J, Liu Z. A discrete-time Geo/G/1 retrial queue with preferred and impatient customers. Appl Math Model 2013; 37: 2552-2561.

[38] Yang T, Li H. The $M/G/1$ retrial queue with the server subject to starting failures. Queueing Syst 1994; 16: 83-96.

[39] Yang T, Posner MJM, Templeton JGC, Li H. An approximation method for the $M/G/1$ retrial queue with general retrial times. Eur J Oper Res 1994; 76: 552-562.

[40] Yang T, Templeton JGC. A survey on retrial queues. Queueing Syst 1987; 2: 201-233.