Yugoslav Journal of Operations Research
16 (2006), Number 1, 45-53

ON THE SINGLE-SERVER RETRIAL QUEUE

Natalia V. DJELLAB

Department of Mathematics,
University of Annaba, Algeria
djellab@yahoo.fr

Received: February 2004 / Accepted: December 2005

Abstract: In this work, we review the stochastic decomposition for the number of customers in M/G/1 retrial queues with reliable server and server subjected to breakdowns which has been the subject of investigation in the literature. Using the decomposition property of M/G/1 retrial queues with breakdowns that holds under exponential assumption for retrial times as an approximation in the non-exponential case, we consider an approximate solution for the steady-state queue size distribution.

Keywords: Retrial queue, stochastic decomposition, breakdown, embedded Markov chain.

1. INTRODUCTION

Retrial queues are characterized by the feature that any arriving customer who finds all servers (and eventually all waiting positions) occupied may repeat his demand after a random amount of time. They are used to solve many practical problems in computer and other communication networks. A review of the main results can be found in [3], [5], [6], [8]. There are different approaches to study retrial queues. We place emphasis on the stochastic decomposition because it leads to simplifications when solving complex models. Stochastic decomposition property takes place in various retrial models, in particular in retrial queuing systems with server vacations [2]; in retrial models with batch arrivals [12] as well as with priority customers [7], [10]. Some applications of the decomposition property for M/G/1 retrial queues have been performed in [4].

In this paper, we review the stochastic decomposition for the number of customers in M/G/1 retrial queues with reliable server and server subjected to breakdowns which has been the subject of investigation in the literature. Assuming the decomposition result for retrial queues with breakdowns and exponential retrial times established in [1] as valid for retrial queues with breakdowns and general retrial times, we consider an approximate solution for the steady-state queue size distribution.

This paper is organized as follows. Model's description is given in the second section. The third section contains a survey of the existing decomposition results. In the fourth section, we consider an approximation method for the computation of the steady-state queue size distribution. In the last section, we show through numerical results how the approximation method works for the M/G/1 retrial queue with breakdowns.

The details of proofs which are available in the literature are omitted, and interested readers are referred back to the original papers.

2. MODELS

We consider single-server queuing systems with no waiting space. Primary customers arrive at the service facility according to a Poisson process with rate $\lambda > 0$. An arriving customer receives immediate service if the server is able to start a service time; otherwise he leaves the service facility temporarily to join the retrial group (orbit). Any orbiting customer persists to ask for service until he gets served. The retrial times are arbitrarily distributed with distribution function $T(x)$ having finite mean $1/\theta$. The service times follow a general distribution with distribution function $B(x)$ having finite mean $1/\gamma$ and Laplace-Stieltjes transform $\tilde{B}(s)$. The input flow of primary customers, service times and retrial times are assumed to be mutually independent.

Model 1: We assume that the server is reliable. Let $C(t)$ be the state of the server at time t. In such a case, $C(t)$ is 0 or 1 depending on whether the server is idle or busy.

Model 2: We assume that the server is subject to Poisson active (when he is busy) and passive (when he is idle) breakdowns with rates μ and η, respectively. The time duration of active and passive interruptions follows random variables D_b and D_i with distribution functions $H(x)$ and $G(x)$ and Laplace-Stieltjes transforms $\tilde{H}(s)$ and $\tilde{G}(s)$, respectively. The variables D_b and D_i hold the assumption of mutual independence. The customers whose service is interrupted by an active breakdown are obliged to leave the service facility either to join the orbit with probability c, or to leave the system with probability $1-c$. The state of the server at time t, $C(t)$, can be 0, 1 or 2 depending on whether the server is, respectively, idle-up, busy or down.

3. STOCHASTIC DECOMPOSITION

In the first time, we review the decomposition results which were established for M/G/1 retrial queues with reliable server (model 1). Consider a non-Markovian process about the number of customers in the system at time t, $\{N(t), t \geq 0\}$. Let $N_o(t)$ be the number of customers in the orbit at time t. Then, $N(t) = C(t) + N_o(t)$. The above process has an embedded Markov chain $\left\{N(\xi_n^+)\right\}$ (ξ_n is the time when the server enters the idle state for the n-th time). We assume that the system is in steady state, which exists if and only if $\rho = \frac{\lambda}{\gamma} < 1$ [6]. From [12], we have that the steady-state distribution of $N(\xi_n^+)$ is also the steady-state distribution of $N(t)$.

Stochastic decomposition for the number of customers in the M/G/1 retrial queue with reliable server was first observed in [12]. Assuming $T(x) = 1 - e^{-\theta x}$, the authors derived the following result about stochastic decomposition for the generating function $Q(z)$ of the steady-state distribution of $\left\{N(\xi_n^+)\right\}$ as $n \to \infty$:

$$Q(z) = \frac{(1-\rho)(1-z)\tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z}\frac{\Phi(z)}{\Phi(1)}, \tag{1}$$

where

$$\Phi(z) = \exp\left\{-\frac{\lambda}{\theta}\int_0^z \frac{1-\tilde{B}(\lambda - \lambda x)}{x - \tilde{B}(\lambda - \lambda x)}dx\right\}.$$

The first factor on the right-hand side of (1) is known as Pollaczek-Khintchine equation for the number of customers in the ordinary M/G/1 queue with infinite waiting space, and the second is the generating function for the number of customers in the retrial queuing system given that the server is idle.

A few years later, Yang and al. [11] demonstrated that the structure of $\Phi(z)/\Phi(1)$ may contain any hint on the structure of the generating function for general retrial time distribution.

Stochastic decomposition becomes more interesting in M/G/1 retrial queues with breakdowns (model 2). Assuming $T(x) = 1 - e^{-\theta x}$, we consider a non-Markovian process about the state of the server, $C(t)$, and the number of customers in the orbit, $N_0(t)$, at time t, $\left\{C(t), N_0(t), t \geq 0\right\}$. The latter has an embedded Markov chain $\left\{N_o(\xi_n^+)\right\}$ (ξ_n is the time when the server enters the idle-up state for the n-th time). From [9], we have that the system is stable if

$$\rho = \lambda\frac{1-\tilde{B}(\mu)}{\mu}\left(1 + \mu\left(E[D_b] + \frac{c}{\lambda}\right)\right) < 1.$$

Aissani and Artalejo [1] introduced an auxiliary queuing system with infinite waiting space, breakdowns of the server and option for leaving the system after an interruption to establish the stochastic decomposition for the generating function $\varphi(z)$ of the steady-state distribution of $\left\{N_o(\xi_n^+)\right\}$ as $n \to \infty$:

$$\varphi(z) = K\frac{\eta z\tilde{G}(\lambda-\lambda z)-(\lambda-\lambda z+\eta)\Omega(z)}{z-\Omega(z)}\frac{P_0(z)}{P_0(1)}, \tag{2}$$

where

$$P_0(z) = K\left[K(1+\eta E[D_i])+(1-\tilde{B}(\mu))(1-\eta K)\left(E[D_b]+\frac{1}{\mu}\right)\right]^{-1}$$
$$\times\exp\left\{\frac{\lambda}{\theta}\int_1^z\frac{1-\Omega(u)}{\Omega(u)-u}du+\frac{\eta}{\theta}\int_1^z\frac{1-\tilde{G}(\lambda-\lambda u)}{\Omega(u)-u}du\right\}$$

with

$$K = \frac{1-\rho}{\lambda\left(1+\eta E[D_i]\right)+\eta\left(1-\rho\right)},$$

$$\Omega(z) = \tilde{B}(\lambda-\lambda z+\mu)+(1-c+cz)\mu\tilde{H}(\lambda-\lambda z)\frac{1-\tilde{B}(\lambda-\lambda z+\mu)}{\lambda-\lambda z+\mu}.$$

The first factor on the right-hand side of (2) is the generating function of the steady-state distribution of the embedded Markov chain at idle-up epochs. It is related to the auxiliary system (without retrials). The remaining factor, $P_0(z)/P_0(1)$, is the generating function for the number of customers in the orbit given that the server is idle-up.

We assume that the decomposition result (2) for exponential retrial times is also valid for general retrial times.

4. APPROXIMATE SOLUTION

In the first time, we introduce some notations. Let ς_n be the time at which the n-th fresh customer arrives at the server; X_i^n be the time elapsed since the last attempt made by the i-th customer in the orbit until instant ξ_n^+. Let $q = \lim_{n\to\infty} N_o(\xi_n^+)$; $X_i = \lim_{n\to\infty} X_i^n$. When $q > 0$, we have a vector $X = (X_1, X_2, ..., X_q)$ of expended retrial times of the q orbiting customers present at an arbitrary time when the server is able to start a new service time (when the server enters the idle state for model 1 et the idle-up state for model 2). We denote by $f_q(x_1, x_2, ..., x_q)$ (or $f_q(x)$) the joint density function of q and X. Define $p_{i,j} = \lim_{t\to\infty} P(C(t) = i; N_o(t) = j)$ and $r_{i,j} = \lim_{n\to\infty} P(C(\varsigma_n^-) = i; N_o(\varsigma_n^-) = j)$ $j = 0,1,2,...;$ $i = 0,1$ (model 1) and $i = 0,1,2$ (model 2).

Now, we consider the decomposition results (1) and (2). One can see that the steady-state distribution of the embedded Markov chains (we denote by $\{d_k, k \ge 0\}$) 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 $d_k = P(q = k)$ for $k \geq 0$. Since Poisson arrivals see time averages, we have $p_{0,k} = r_{0,k}$ for $k \geq 0$. Suppose that there are $k > 0$ customers in the orbit at an arbitrary time when the server is able to start a new service time. In such a case, we have

$$d_k = \int_0^\infty f_k(x)dx; \qquad r_{0,k} = \int_0^\infty P(\delta(k;x) = 0)\, f_k(x)dx;$$

where $\delta(k;x) = 0$ if the next served customer is not one of the k orbiting customers, otherwise $\delta(k;x) = 1$. Since expended retrial times $X_1, X_2, ..., X_k$ of the k 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 [11] for retrial models with reliable server: $f_k(x) \approx d_k \theta^k \prod_{i=1}^{k} \left(1 - T(x_i)\right)$. It is based on the intuitive consideration that the mean retrial time is very small relative to the mean service time (for retrial models with breakdowns, we add that the mean retrial time is also very small relative to the mean time duration of interruptions).

Using the above approximation, it was established that $r_{0,k} \approx d_k b_k$, where $b_k = \int_0^\infty \left(1 - m(t)\right)^k \lambda e^{-\lambda t} dt$ with $m(t) = \int_0^t \theta\left(1 - T(u)\right) du.$

We assume that $\{a_k, k \geq 0\}$ is already known. Under this assumption, we can express the results (1) and (2) in the following common form:

$$d_k = \frac{1}{1-\rho} \sum_{i=0}^{k} a_i r_{0,k-i} \tag{3}$$

with

$$r_{0,k} \approx d_k b_k, \tag{4}$$

$$\sum_{k=0}^{\infty} d_k = 1. \tag{5}$$

The set of equations (3)-(5) gives an approximate solution to $\{r_{0,k}, k \geq 0\}$ and $\{d_k, k \geq 0\}$. Let $\{\hat{r}_{0,k}, k \geq 0\}$ and $\{\hat{d}_k, k \geq 0\}$ be the approximations to $\{r_{0,k}, k \geq 0\}$ and $\{d_k, k \geq 0\}$, respectively. From (3)-(5), it is easy to find the following computational procedure:

$$\hat{d}_k = g_k \hat{d}_0, \qquad k = 0,1,2...; \tag{6}$$

$$\hat{r}_{0,k} = g_k b_k \hat{d}_0, \qquad k = 0,1,2,...; \tag{7}$$

$$\hat{d}_0 = \frac{1}{\sum_{k=0}^{\infty} g_k}; \tag{8}$$

where

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

Once system steady-state probabilities are evaluated, various performance measures can be calculated. Let N be the number of customers in the system at an arbitrary time when the server is able to start a new service time. Then, we have

$$E[N] \approx \sum_{k=0}^{\infty} k\hat{d}_k; \qquad Var[N] \approx \sum_{k=0}^{\infty} k^2 \hat{d}_k - (E[N])^2 .$$

Let N_o be the number of customers in the orbit at an arbitrary time. Then, we have (for models without breakdowns)

$$E[N_o] = E[N] - \rho; \quad Var[N_o] = \sum_{k=0}^{\infty} k^2 \left(\hat{r}_{0,k} + \hat{d}_{k+1} - \hat{r}_{0,k+1}\right) - (E[N_o])^2 .$$

5. NUMERICAL RESULTS

In this section, we examine the performance of the approximation discussed in the preceding section in the case of M/G/1 retrial queue subjected to breakdowns. We consider the following service and retrial time distributions: exponential (E), two-stage Erlang (E_2), and two-stage hyper exponential (H_2). Throughout this section, we let the mean service time $1/\gamma$ be a unit time, the rate of active breakdowns μ as well as the rate of passive breakdowns η be 0.02, the mean time duration of active interruption $E[D_b]$ as well as the mean time duration of passive interruption $E[D_i]$ be 0.2, the recovery factor c be 0.9.

Tables 1, 2 and 3 present the approximation outcomes calculated according to (6)-(8) against those from a simulation study (at 95-percent confidence intervals) for the M/M/1, M/E_2/1 and M/H_2/1 retrial models with breakdowns, respectively. From the simulation results given in tables 1, 2 and 3, we can see that the mean system size at an arbitrary idle-up epoch $E[N]$ is an increasing function of the second moments of both the service time distribution and the retrial time distribution. This property also takes place in the approximate results.

Concerning M/G/1 retrial queues with reliable server, 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) [11]. In tables 1, 2 and 3 we observe that the approximate results are close to the simulation ones when there are M/M/1 and M/E_2/1 retrial models with breakdowns for which the coefficient of variation of service times $cs \le 1$. In the case of M/H_2/1 retrial model with $cs = 2$, the approximation method works well as long as the traffic intensity ρ is relatively low (see $\lambda \in \{0.1; 0.3\}$). On the other hand, the approximation fails when the traffic intensity is high (see $\lambda = 0.6*$): the difference between the two solutions is highly significant.

Table 1: The M/M/1 retrial model with breakdowns (*cs*=1)

λ	θ	Retrial times					
		E		E_2		H_2	
		cv=1		*cv*≈0.7		*cv*=1.5	
		E[N] approx	E[N] simul	E[N] approx	E[N] simul	E[N] approx	E[N] simul
0.1	1	0.1262	0.1253	0.1234	0.1224	0.1320	0.1303
	3.3	0.1181	0.1179	0.1173	0.1166	0.1201	0.1191
	10	0.1158	0.1157	0.1155	0.1150	0.1165	0.1158
0.3	1	0.5684	0.5710	0.5410	0.5501	0.6126*	0.6771*
	3.3	0.4765	0.4761	0.4677	0.4656	0.4943	0.5004
	10	0.4498	0.4495	0.4468	0.4457	0.4562	0.4582
0.6	1	2.4414	2.4725	2.3006	2.3493	2.5707*	3.2444*
	3.3	1.7988	1.7947	1.7492	1.7411	1.8710*	2.0144*
	10	1.6117	1.6082	1.5945	1.5884	1.6414	1.6510

Table 2: The M/E_2/1 retrial model with breakdowns (*cs*≈0.7)

λ	θ	Retrial times					
		E		E_2		H_2	
		cv=1		*cv*≈0.7		*cv*=1.5	
		E[N] approx	E[N] simul	E[N] approx	E[N] simul	E[N] approx	E[N] simul
0.1	1	0.1241	0.1232	0.1213	0.1200	0.1300	0.1283
	3.3	0.1160	0.1153	0.1151	0.1141	0.1180	0.1165
	10	0.1136	0.1134	0.1134	0.1129	0.1143	0.1137
0.3	1	0.5399	0.5480	0.5118	0.5211	0.5859*	0.6528*
	3.3	0.4474	0.4454	0.4383	0.4344	0.4659	0.4740
	10	0.4204	0.4198	0.4174	0.4143	0.4272	0.4302
0.6	1	2.2434*	2.3065*	2.0957*	2.1656*	2.3925*	3.1457*
	3.3	1.5927	1.5872	1.5401	1.5303	1.6733*	1.8225*
	10	1.4031	1.3976	1.3849	1.3781	1.4364	1.4475

According to the idea of the approximation, the retrial intensity and the type of retrial time distribution (or its coefficient of variation *cv*) seem to be the important factors affecting its performance. Regarding M/G/1 retrial queues without breakdowns, it was shown that 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 is fairly close to that of the exponential distribution ($cv < 4$) [11]. We have examined the effects of the retrial intensity and those of the type of retrial time distribution on the performance of the approximation in the case of M/G/1 retrial queues with breakdowns. From numerical results shown in tables 1, 2, 3 and also 4, we can see that the approximation deteriorates as the mean retrial time $1/\theta$ approaches the mean service time $1/\gamma$. It fails when $1/\theta$ is not sufficiently small relative to $1/\gamma$ and the traffic intensity is relatively high (the failure of the approximation is denoted by *). We can also see that the accuracy of the approximation deteriorates as the retrial time distribution departs from the exponential distribution in the sense that its coefficient of variation $cv > 1$. Further, we have observed that the approximate results are close to the simulation ones, when the coefficient of variation of retrial times $cv < 3$ (for example, see table 4).

Table 3: The M/H_2/1 retrial model with breakdowns (*cs*=2)

λ	θ	Retrial times					
		E		E_2		H_2	
		cv=1		*cv*≈0.7		*cv*=1.5	
		E[N] approx	E[N] simul	E[N] approx	E[N] simul	E[N] approx	E[N] simul
0.1	1	0.1326	0.1342	0.1301	0.1337	0.1378	0.1419
	3.3	0.1250	0.1264	0.1242	0.1261	0.1268	0.1293
	10	0.1228	0.1237	0.1225	0.1236	0.1234	0.1245
0.3	1	0.6813	0.6929	0.6574	0.6724	0.7178*	0.7971*
	3.3	0.5960	0.5998	0.5884	0.5938	0.6106	0.6227
	10	0.5712	0.5743	0.5686	0.5720	0.5765	0.5830
0.6*	1	3.1686	3.4969	3.0530	3.3042	3.1990	3.7919
	3.3	2.6031	2.8536	2.5635	2.8139	2.6379	2.9054
	10	2.4384	2.7224	2.4248	2.7035	2.4542	2.8039
	50	2.3735	2.6733	2.3708	2.6234	2.3771	2.7232

Table 4: The M/M/1 retrial model with breakdowns

λ	θ	Retrial times			
		H_2		H_2	
		cv=2		*cv*=3	
		E[N] approx	E[N] simul	E[N] approx	E[N] simul
0.3	3.3	0.5154*	0.5804*	0.5631*	0.7963*
	10	0.4649	0.4714	0.4870*	0.5635*
	20	0.4510	0.4543	0.4632	0.4909*
	30	0.4462	0.4476	0.4545	0.4666
	50	0.4424	0.4414	0.4476	0.4496
0.6	10	1.6791*	1.8045*	1.7656*	2.4285*
	20	1.6023	1.6340	1.6549*	1.9077*
	30	1.5749	1.5949	1.6123*	1.8109*
	50	1.5535	1.5564	1.5775*	1.6628*

We conclude that increasing the traffic intensity and increasing the coefficient of variation of service time distribution as well as the coefficient of variation of retrial time distribution have an adverse influence on the performance of the approximation.

REFERENCES

[1] Aissani, A., and Artalejo, J.R., "On the single server retrial queue subject to breakdowns", *Queueing Systems,* 30 (1998) 309-321.

[2] Artalejo, J.R., "Analysis of an M/G/1 queue with constant repeated attempts and server vacations", *Computer Ops Res.,* 24 (1997) 493-504.

[3] Artalejo, J.R., "Retrial queues with a finite number of sources", *Journal of Korean Mathematical Society*, 35 (1998) 503-525.

[4] Artalejo, J.R., and Falin, G., "Stochastic decomposition for retrial queues", *TOP,* 2 (1994), 3289-3342.

[5] Choi, B.D., and Chang, Y., "Single server retrial queues with priority calls", *Mathematical and Computer Modelling,* 30 (1999) 7-32.

[6] Falin, G., "A survey of retrial queues", *Queueing Systems,* 7 (1990), 127-168.

[7] Falin, G., Artalejo, J.R., and Martin, M., "On the single server retrial queue with priority customers", *Queueing Systems,* 14 (1993) 439-455.

[8] Falin, G., and Templeton, J.G.C., *Retrial Queues*, Chapman and Hall, London, 1997.

[9] Kulkarni, V.G., and Choi, B.D., "Retrial queues with server subject to breakdowns and repairs", *Queueing Systems,* 7 (1990), 191-208.

[10] Li, H., and Yang, T., "Geo/G/1 discrete time retrial queue with Bernoulli schedule", *European Journal of Operational Research,* 111 (1998), 629-649.

[11] Yang, T., and al., "An approximation method for M/G/1 retrial queue with general retrial times", *European Journal of Operational Research,* 76 (1994), 552-562.

[12] Yang, T., and Templeton, J.G.C., "A survey on retrial queues", *Queueing Systems,* 2 (1987), 201-233.