

M/M/1 RETRIAL QUEUES WITH CONSTANT RETRIAL POLICY

Abdul Rof V

Assistant Professor, Department of Mathematics, KAHM Unity Women's College Manjeri, India

***Corresponding Author:**

***Abdul Rof V**

Abstract:

An M/M/1 retrial queue with constant retrial policy is discussed. The system is modelled as a level-independent quasi-birth-and-death (QBD) process and its infinitesimal generator matrix is obtained. The stability condition $\lambda\lambda+\mu<\mu\nu$ is derived from the mean drift condition. The rate matrix R of the matrix-geometric method is obtained in closed form, and from it the limiting probabilities of the system are calculated explicitly. Several performance measures, such as the mean orbit size, the mean system size, the mean waiting times and the expected number of retrials, are derived. The analytical results are verified numerically and by simulation, and the effect of the system parameters is illustrated through tables and graphs.

Keywords: *Retrial queue, orbit, infinitesimal generator matrix, constant retrial rate, matrix-geometric method, quasi-birth-and-death process*

Introduction

If a person obtains a busy signal when trying to call another over the phone, he repeats his attempt until the connection is made, or until the expiration of a specified time. If the line is busy, he cannot wait for service; so he puts down the phone and tries again later. Thus the flow of calls in a telephone network consists of two parts: (i) the flow of primary calls, and (ii) the flow of repeated calls. A customer who fails in his first attempt becomes a source of repeated trials. Queues of this type are termed retrial queues. They have applications in telephone and communication networks, the internet, aircraft landing, and everyday queues such as that of shoppers who leave a long line and return after some time, when the queue may be shorter.

Retrial queues have been studied extensively; comprehensive accounts are available in the surveys of Falin [1] and Kim and Kim [2], the monographs of Falin and Templeton [3] and Artalejo and Gómez-Corral [4], and the classified bibliographies of Artalejo [5, 6]. A comparison between standard and retrial queueing systems is given in [7].

In the classical retrial policy each customer in the orbit retries independently of the others, so that the total retrial rate is $n\mu$ when there are n customers in the orbit. In the constant retrial policy the retrial rate is μ whenever the orbit is non-empty, independently of the number of customers in it. This policy was introduced by Fayolle [8] in the study of a telephone exchange in which only the customer at the head of the orbit can retry, and was studied by Farahmand [9] under the name “queue with repeated demands”. It arises naturally in random-access protocols such as CSMA/CD [10], and it has been analysed for retrial control policies [11], for the M/M/2 queue [12], for M/M/r/r + d queues [13], for single-server queues with general retrial times [14, 15], and for multiserver queues [16].

Matrix-analytic methods [17, 18, 19] are a natural tool for retrial queues; see the bibliographical guide of Gómez-Corral [20]. Under the classical policy the underlying Markov chain is level-dependent, and generating functions are usually employed. Under the constant policy the chain is a level-independent QBD process, so the matrix-geometric method applies directly.

Consider a single server queueing system in which customers arrive according to a Poisson process with rate λ . These customers are called primary customers. If the server is idle at the arrival of a primary customer, the arriving customer begins to be served immediately and leaves the system after service completion. Otherwise, if the server is busy, the arriving customer leaves the service area and joins a pool of unsatisfied customers called the orbit. The orbit as a whole generates retrials according to a Poisson process with intensity μ whenever it is non-empty. The service times of both primary and retrial customers are exponentially distributed with rate ν .

At time t , let $N(t)$ be the number of customers in the orbit and $C(t)$ the server status,

$$C(t) = \begin{cases} 1, & \text{if the server is busy,} \\ 0, & \text{if the server is idle.} \end{cases}$$

The process $(C(t), N(t))$ describes the number of customers in the system in the simplest way, and it is the most important process associated with this queueing system. When the service time distribution is not exponential, this process is not Markov.

In this paper we consider the constant retrial policy, in which retrial is level independent, i.e. the retrial rate does not depend on the number of customers in the orbit. We find the infinitesimal generator matrix, obtain the rate matrix R explicitly, and compute the limiting probabilities and the main performance measures from it.

2. Model Description

The queueing process evolves in the following manner. Suppose that the $(i-1)$ th customer completes his service at the epoch t_{i-1} and the server becomes free. Even if there are customers in the orbit who want service, they cannot occupy the server immediately, because they do not know the server state. So the next (i) th customer enters the server only after a time interval R_i during which the server is idle. If the number of orbital customers at the epoch t_{i-1} is n , the random variable R_i has an exponential distribution with parameter $\lambda + \mu$ if $n \geq 1$, and parameter λ if $n = 0$. When $n \geq 1$, the i th customer is a primary customer with probability $\frac{\lambda}{\lambda + \mu}$ and an orbital customer with probability $\frac{\mu}{\lambda + \mu}$. At the epoch $\xi_i = t_{i-1} + R_i$, the i th customer starts his service, which continues for a time $S_i \sim \exp(\nu)$. All primary customers arriving during the service time join the orbit; retrials made during this time are unsuccessful and do not change the state of the system. At the epoch $t_i = \xi_i + S_i$ the i th customer completes his service and the server becomes free again.

Since the inter-arrival, retrial and service times are exponential, the process $\{(C(t), N(t)), t \geq 0\}$ is a continuous-time Markov chain with state space $\{0,1\} \times \{0,1,2, \dots\}$. We call $\{(0,n), (1,n)\}$ the n th level.

3. Infinitesimal Generator Matrix

The transitions of the Markov chain $\{(C(t), N(t)), t \in \mathbb{R}_+\}$ are as follows. From $(0,n)$: a primary arrival leads to $(1,n)$ at rate λ , and, if $n \geq 1$, a successful retrial leads to $(1,n-1)$ at rate μ . From $(1,n)$: a service completion leads to $(0,n)$ at rate ν , and a primary arrival joins the orbit, leading to $(1,n+1)$ at rate λ . Hence the generator, with states ordered $(0,0), (1,0), (0,1), (1,1), \dots$, is

	(0,0)	(1,0)	(0,1)	(1,1)	(0,2)	(1,2)	(0,3)	...
(0,0)	$-\lambda$	λ	0	0	0	0	0	
(1,0)	ν	$-(\lambda + \nu)$	0	λ	0	0	0	
(0,1)	0	μ	$-(\lambda + \mu)$	λ	0	0	0	
(1,1)	0	0	ν	$-(\lambda + \nu)$	0	λ	0	
(0,2)	0	0	0	μ	$-(\lambda + \mu)$	λ	0	
(1,2)	0	0	0	0	ν	$-(\lambda + \nu)$	0	...

In block form,

$$Q = \begin{pmatrix} B_1 & A_0 & 0 & 0 & \dots \\ A_2 & A_1 & A_0 & 0 & \dots \\ 0 & A_2 & A_1 & A_0 & \dots \\ \vdots & & \ddots & \ddots & \ddots \end{pmatrix}, \quad (1)$$

where

$$B_1 = \begin{pmatrix} -\lambda & \lambda \\ \nu & -(\lambda + \nu) \end{pmatrix}, \quad A_0 = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}, \quad A_1 = \begin{pmatrix} -(\lambda + \mu) & \lambda \\ \nu & -(\lambda + \nu) \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & \mu \\ 0 & 0 \end{pmatrix}.$$

Since the blocks do not depend on the level n , Q is the generator of a level-independent QBD process [17, 18].

4. Stability Condition

Let $\pi = (\pi_0, \pi_1)$ be the stationary probability vector of the generator

$$A = A_0 + A_1 + A_2 = \begin{pmatrix} -(\lambda + \mu) & \lambda + \mu \\ \nu & -\nu \end{pmatrix},$$

which describes the server status when the orbit is far from empty. Then $\pi A = 0$ and $\pi e = 1$, where e is a column vector of ones; that is,

$$-\pi_0(\lambda + \mu) + \pi_1\nu = 0, \quad \pi_0 + \pi_1 = 1.$$

Hence $\pi_1 = \frac{\lambda + \mu}{\nu} \pi_0$, so that $\pi_0 \left[\frac{\lambda + \mu}{\nu} + 1 \right] = 1$, which gives

$$\pi_0 = \frac{\nu}{\lambda + \mu + \nu}, \quad \pi_1 = \frac{\lambda + \mu}{\lambda + \mu + \nu}.$$

By the mean drift condition [17, 18], the QBD process is positive recurrent if and only if $\pi A_0 e < \pi A_2 e$, i.e.

$$(\pi_0, \pi_1) \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} < (\pi_0, \pi_1) \begin{pmatrix} 0 & \mu \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

so that $\lambda\pi_1 < \mu\pi_0$, i.e. $\lambda \frac{\lambda + \mu}{\lambda + \mu + \nu} < \mu \frac{\nu}{\lambda + \mu + \nu}$. Therefore the system is stable if and only if

$$\lambda(\lambda + \mu) < \mu\nu, \quad \text{equivalently} \quad \sigma := \frac{\lambda(\lambda + \mu)}{\mu\nu} < 1. \quad (2)$$

Writing $\rho = \lambda/\nu$, the condition is $\rho < \frac{\mu}{\lambda + \mu}$, which agrees with the result of Fayolle [8].

Remark 1. Since $\frac{\mu}{\lambda + \mu} < 1$, condition (2) is stronger than the condition $\rho < 1$ of the standard M/M/1 queue: under the constant policy the server may remain idle while customers wait in the orbit. As $\mu \rightarrow \infty$ the condition reduces to $\rho < 1$.

5. System State Probabilities

Assume that (2) holds. Let $x = (x_0, x_1, x_2, \dots)$ be the steady state probability vector of the process $\{(C(t), N(t)): t \geq 0\}$, where $x_n = (x_n(0), x_n(1))$, $n = 0, 1, 2, \dots$, and $x_n(c) = \lim_{t \rightarrow \infty} P\{C(t) = c, N(t) = n\}$. Then $xQ = 0$ and $xe = 1$, so that

$$x_0 B_1 + x_1 A_2 = 0, \quad (3)$$

$$x_{n-1} A_0 + x_n A_1 + x_{n+1} A_2 = 0, \quad n \geq 1. \quad (4)$$

By the matrix-geometric theorem [17], $x_n = x_0 R^n$, $n \geq 0$, where R is the minimal non-negative solution of the matrix quadratic equation

$$A_0 + RA_1 + R^2 A_2 = 0. \quad (5)$$

Theorem 1. Under condition (2), the rate matrix is $R = \begin{pmatrix} 0 & 0 \\ \frac{\lambda}{\mu} & \sigma \end{pmatrix}$, $\sigma = \frac{\lambda(\lambda+\mu)}{\mu v}$, and $R^n = \sigma^{n-1} R$ for $n \geq 1$. The spectral radius of R is $\sigma < 1$.

Proof. The first row of A_0 is zero, which corresponds to the fact that the level can increase only from a state in which the server is busy. Hence the first row of R is zero [17], and we may write $R = \begin{pmatrix} 0 & 0 \\ a & b \end{pmatrix}$ with $a, b \geq 0$. Then $R^2 = \begin{pmatrix} 0 & 0 \\ ab & b^2 \end{pmatrix} = bR$, and the second row of (5) gives

$$\begin{aligned} -a(\lambda + \mu) + bv &= 0, \\ \lambda + a\lambda - b(\lambda + v) + ab\mu &= 0. \end{aligned}$$

The first equation gives $a = \frac{bv}{\lambda + \mu}$. Substituting in the second equation and multiplying by $\lambda + \mu$,

$$\mu v b^2 - [\lambda(\lambda + \mu) + \mu v]b + \lambda(\lambda + \mu) = 0,$$

that is, $(b - 1)(\mu v b - \lambda(\lambda + \mu)) = 0$. The two roots are $b = 1$ and $b = \sigma$, and the minimal non-negative solution is $b = \sigma$, since $\sigma < 1$. Then $a = \frac{\sigma v}{\lambda + \mu} = \frac{\lambda}{\mu}$. Since $R^2 = \sigma R$, induction gives $R^n = \sigma^{n-1} R$. The eigenvalues of the triangular matrix R are 0 and σ . $\square$

Theorem 2. Under condition (2), the limiting probabilities are

$$\begin{aligned} x_0 &= (x_0(0), x_0(1)) = (1 - \sigma)(1, \rho), \\ x_n &= (x_n(0), x_n(1)) = (1 - \sigma)\sigma^n \left(\frac{\lambda}{\lambda + \mu}, \rho \right), \quad n \geq 1, \end{aligned}$$

where $\rho = \lambda/v$ and $\sigma = \lambda(\lambda + \mu)/(\mu v)$.

Proof. The boundary equation (3) with $x_1 = x_0 R$ becomes $x_0(B_1 + RA_2) = 0$. Here $RA_2 = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}$, so

$$B_1 + RA_2 = \begin{pmatrix} -\lambda & \lambda \\ v & -v \end{pmatrix},$$

which gives $-\lambda x_0(0) + v x_0(1) = 0$, i.e. $x_0(1) = \rho x_0(0)$. The normalising condition is $\sum_{n \geq 0} x_0 R^n e = x_0(I - R)^{-1} e = 1$. Now

$$(I - R)^{-1} = \begin{pmatrix} 1 & 0 \\ \frac{\lambda}{\mu(1 - \sigma)} & \frac{1}{1 - \sigma} \end{pmatrix}, \quad (I - R)^{-1} e = \begin{pmatrix} 1 \\ \frac{\lambda + \mu}{\mu(1 - \sigma)} \end{pmatrix},$$

so that $x_0(0) \left[1 + \frac{\rho(\lambda + \mu)}{\mu(1 - \sigma)} \right] = x_0(0) \left[1 + \frac{\sigma}{1 - \sigma} \right] = \frac{x_0(0)}{1 - \sigma} = 1$. Hence $x_0(0) = 1 - \sigma$ and $x_0(1) = (1 - \sigma)\rho$. Finally, for $n \geq 1$,

Theorem 1 gives $x_n = x_0 R^n = \sigma^{n-1} x_0 R = \sigma^{n-1} x_0(1) \begin{pmatrix} \lambda \\ \mu \end{pmatrix}$, and $\rho \frac{\lambda}{\mu} = \frac{\lambda}{\lambda + \mu} \sigma$, which completes the proof. $\square$

Corollary 1.

The probability that the server is busy is $\sum_{n \geq 0} x_n(1) = \rho$, and the probability that it is idle is $1 - \rho$.

Given that the server is busy, the orbit size has the geometric distribution $P\{N = n | C = 1\} = (1 - \sigma)\sigma^n$, $n \geq 0$.

The distribution of the orbit size is $P\{N = 0\} = (1 - \sigma)(1 + \rho)$, $P\{N = n\} = (1 - \sigma)\theta \sigma^n$, $n \geq 1$,

$$\text{where } \theta = \frac{\lambda}{\lambda + \mu} + \rho = \frac{\lambda(\lambda + \mu + v)}{v(\lambda + \mu)}.$$

6. Performance Measures

Using Theorem 2, the main performance measures of the system are as follows.

Server utilisation: $P\{C = 1\} = \rho = \lambda/v$. By the PASTA property [21], this is also the probability that a primary customer is blocked and joins the orbit.

Probability that the system is empty: $x_0(0) = 1 - \sigma$.

Probability that the orbit is empty: $P_0 = (1 - \sigma)(1 + \rho)$.

Mean number of customers in the orbit:

$$L_o = \sum_{n \geq 1} n (x_n(0) + x_n(1)) = \frac{\theta \sigma}{1 - \sigma} = \frac{\lambda^2 (\lambda + \mu + \nu)}{\nu (\mu \nu - \lambda (\lambda + \mu))}.$$

Mean number of customers in the system: $L = L_o + \rho$.

Mean waiting time in the orbit and mean sojourn time (by Little's law [21]): $W_o = L_o/\lambda$ and $W = W_o + 1/\nu$.

Expected number of retrials per customer. The orbit generates retrials at rate μ whenever it is non-empty; those made while the server is busy are unsuccessful. Hence the mean number of retrials made per customer is

$$\bar{r} = \frac{\mu(1 - P_0)}{\lambda} = \frac{\mu}{\lambda} [1 - (1 - \sigma)(1 + \rho)].$$

The rate of successful retrials is $\mu \sum_{n \geq 1} x_n(0) = \mu(\sigma - \rho) = \lambda \rho$, which equals the rate at which customers enter the orbit, as it must.

Remark 2 (The limiting case $\mu \rightarrow \infty$). As $\mu \rightarrow \infty$, $\sigma \rightarrow \rho$, $\theta \rightarrow \rho$, and $L_o \rightarrow \frac{\rho^2}{1-\rho}$, which is the mean queue length L_q of the standard M/M/1 queue [21]. This is expected, since with infinitely fast retrials a customer from the orbit occupies the server as soon as it becomes free.

7. Numerical Illustration

In this section the results are illustrated numerically. Throughout, the service rate is taken as $\nu = 1$.

7.0.0.1 Verification.

The closed-form results were checked in three independent ways: (a) the matrix R was computed by the iteration $R \leftarrow (A_0 + R^2 A_2) A_1^{-1}$ and agreed with Theorem 1 to within 5×10^{-15} ; (b) the generator was truncated at 600 orbit levels and the linear system $xQ = 0$, $x_e = 1$ was solved directly; (c) the system was simulated for 2×10^6 time units. Table 1 shows the agreement (Exact: Theorem 2; CTMC: truncated generator; Sim.: simulation).

Comparison of closed-form, numerical and simulation results ($\nu = 1$).

λ	μ	L_o exact	L_o CTMC	L_o sim.	$x_0(0)$ exact	$x_0(0)$ CTMC	$P\{C=0\}$ exact	$P\{C=0\}$ sim.
0.3	1	0.3393	0.3393	0.3379	0.6100	0.6100	0.7000	0.6992
0.5	2	1.1667	1.1667	1.1544	0.3750	0.3750	0.5000	0.5014
0.6	5	1.4488	1.4488	1.4576	0.3280	0.3280	0.4000	0.3996
0.4	0.5	2.1714	2.1714	2.1916	0.2800	0.2800	0.6000	0.5996

7.0.0.2 Effect of the arrival rate.

Table 2 gives the performance measures for $\mu = 2$ and varying λ . Here the stability condition is $\lambda < \sqrt{3} - 1 \approx 0.732$. As λ increases, the probability of an empty system decreases, while L_o , L , W_o and W increase, slowly at first and very sharply as λ approaches the stability limit (Figure 1).

Performance measures for varying λ ($\mu = 2$, $\nu = 1$).

λ	σ	$x_0(0)$	P_0	L_o	L	W_o	W	$\bar{r}$
0.1	0.1050	0.8950	0.9845	0.0173	0.1173	0.1732	1.1732	0.3100
0.2	0.2200	0.7800	0.9360	0.0821	0.2821	0.4103	1.4103	0.6400
0.3	0.3450	0.6550	0.8515	0.2267	0.5267	0.7557	1.7557	0.9900
0.4	0.4800	0.5200	0.7280	0.5231	0.9231	1.3077	2.3077	1.3600
0.5	0.6250	0.3750	0.5625	1.1667	1.6667	2.3333	3.3333	1.7500
0.6	0.7800	0.2200	0.3520	2.9455	3.5455	4.9091	5.9091	2.1600
0.7	0.9450	0.0550	0.0935	16.4818	17.1818	23.5455	24.5455	2.5900

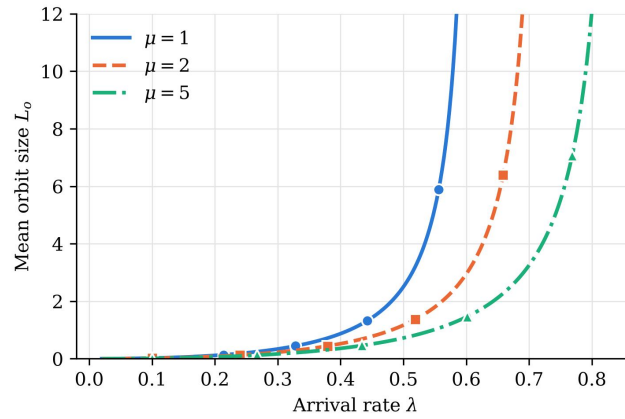
7.0.0.3 Effect of the retrial rate.

Table 3 gives the performance measures for $\lambda = 0.4$ and varying μ ; the system is stable for $\mu > 4/15 \approx 0.267$. As μ increases, L_o , L , W_o and W decrease and approach the M/M/1 values ($L_q = 0.2667$, $L = 0.6667$, $W_q = 0.6667$, $W = 1.6667$), while the expected number of retrials per customer increases, because a fast-retrying orbit also makes many unsuccessful attempts during busy periods (Figure 2). Figure 3 shows the orbit size distribution: smaller retrial rates give a heavier tail.

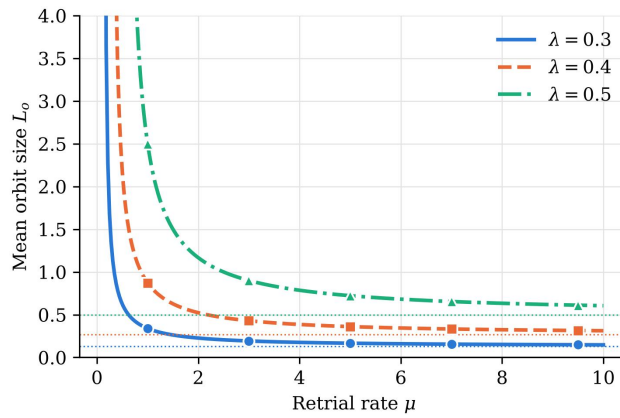
Performance measures for varying μ ($\lambda = 0.4$, $\nu = 1$).

μ	σ	$x_0(0)$	P_0	L_o	L	W_o	W	$\bar{r}$
0.5	0.7200	0.2800	0.3920	2.1714	2.5714	5.4286	6.4286	0.7600
1	0.5600	0.4400	0.6160	0.8727	1.2727	2.1818	3.1818	0.9600
2	0.4800	0.5200	0.7280	0.5231	0.9231	1.3077	2.3077	1.3600
3	0.4533	0.5467	0.7653	0.4293	0.8293	1.0732	2.0732	1.7600
5	0.4320	0.5680	0.7952	0.3606	0.7606	0.9014	1.9014	2.5600

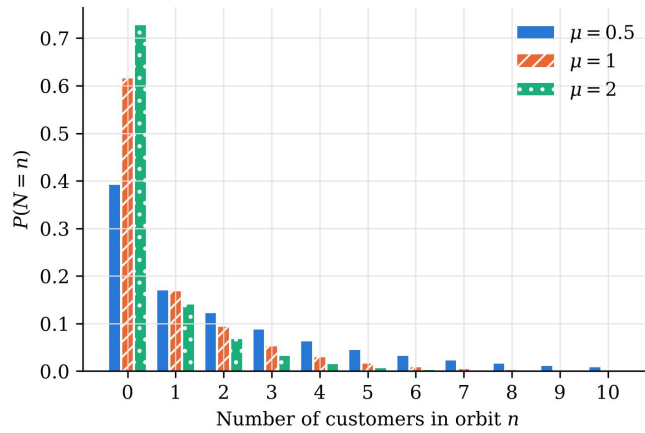
μ	σ	$x_0(0)$	P_0	L_0	L	W_0	W	$\bar{r}$
10	0.4160	0.5840	0.8176	0.3123	0.7123	0.7808	1.7808	4.5600
50	0.4032	0.5968	0.8355	0.2756	0.6756	0.6890	1.6890	20.5600



Mean orbit size L_0 against the arrival rate λ for $\mu = 1, 2, 5$ ($v = 1$).



Mean orbit size L_0 against the retrial rate μ for $\lambda = 0.3, 0.4, 0.5$ ($v = 1$). The dotted lines show the M/M/1 limit $\rho^2/(1 - \rho)$.



Orbit size distribution $P\{N = n\}$ for $\lambda = 0.4$, $v = 1$ and $\mu = 0.5, 1, 2$.

8. Conclusion

The M/M/1 retrial queue with constant retrial policy has been analysed as a level-independent QBD process. The stability condition $\lambda(\lambda + \mu) < \mu v$ was obtained from the mean drift condition, and the rate matrix R was found in closed form. Since R has rank one, the limiting probabilities are obtained explicitly: given that the server is busy, the orbit size is geometric with parameter $\sigma = \lambda(\lambda + \mu)/(\mu v)$. Explicit expressions for the mean orbit size, the mean system size, the mean waiting times and the expected number of retrials were derived and verified numerically and by simulation.

M/M/1 retrial queues with the classical retrial policy can be solved using generating functions; the solutions of various retrial queues by generating functions are discussed in [3]. In that case the generator is level-dependent and the matrix-geometric method is rather difficult to apply. Under the constant retrial policy, in contrast, the matrix-geometric method gives the solution directly and in closed form. Extensions to phase-type service times, batch arrivals and multiple servers

[16] can be treated along the same lines, with R computed numerically.

References

1. G. I. Falin, A survey of retrial queues, *Queueing Systems* 7 (1990) 127–167.
2. J. Kim and B. Kim, A survey of retrial queueing systems, *Annals of Operations Research* 247 (2016) 3–36.
3. G. I. Falin and J. G. C. Templeton, *Retrial Queues*, Chapman & Hall, London, 1997.
4. J. R. Artalejo and A. Gómez-Corral, *Retrial Queueing Systems: A Computational Approach*, Springer, Berlin, 2008.
5. J. R. Artalejo, A classified bibliography of research on retrial queues: Progress in 1990–1999, *Top* 7 (1999) 187–211.
6. J. R. Artalejo, Accessible bibliography on retrial queues: Progress in 2000–2009, *Mathematical and Computer Modelling* 51 (2010) 1071–1081.
7. J. R. Artalejo and G. Falin, Standard and retrial queueing systems: a comparative analysis, *Revista Matemática Complutense* 15 (2002) 101–129.
8. G. Fayolle, A simple telephone exchange with delayed feedbacks, in: O. J. Boxma, J. W. Cohen and H. C. Tijms (eds.), *Teletraffic Analysis and Computer Performance Evaluation*, Elsevier Science Publishers B.V., Amsterdam, 1986, pp. 245–253.
9. K. Farahmand, Single line queue with repeated demands, *Queueing Systems* 6 (1990) 223–228.
10. B. D. Choi, Y. W. Shin and W. C. Ahn, Retrial queues with collision arising from unslotted CSMA/CD protocol, *Queueing Systems* 11 (1992) 335–356.
11. B. D. Choi, K. K. Park and C. E. M. Pearce, An M/M/1 retrial queue with control policy and general retrial times, *Queueing Systems* 14 (1993) 275–292.
12. J. R. Artalejo, Stationary analysis of the characteristics of the M/M/2 queue with constant repeated attempts, *Opsearch* 33 (1996) 83–95.
13. M. F. Ramalhoto and A. Gómez-Corral, Some decomposition formulae for M/M/r/r+d queues with constant retrial rate, *Stochastic Models* 14 (1998) 123–145.
14. A. Gómez-Corral, Stochastic analysis of a single server retrial queue with general retrial times, *Naval Research Logistics* 46 (1999) 561–581.
15. I. Atencia and P. Moreno, A single-server retrial queue with general retrial times and Bernoulli schedule, *Applied Mathematics and Computation* 162 (2005) 855–880.
16. J. R. Artalejo, A. Gómez-Corral and M. F. Neuts, Analysis of multiserver queues with constant retrial rate, *European Journal of Operational Research* 135 (2001) 569–581.
17. M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins University Press, Baltimore, 1981.
18. G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, 1999.
19. Q.-M. He, *Fundamentals of Matrix-Analytic Methods*, Springer, New York, 2014.
20. A. Gómez-Corral, A bibliographical guide to the analysis of retrial queues through matrix analytic techniques, *Annals of Operations Research* 141 (2006) 163–191.
21. D. Gross, J. F. Shortle, J. M. Thompson and C. M. Harris, *Fundamentals of Queueing Theory*, 4th ed., John Wiley & Sons, Hoboken, NJ, 2008.