Air Force Institute of Technology

AFIT Scholar

Faculty Publications

11-2006

An M/M/1 Retrial Queue with Unreliable Server

Nathan P. Sherman

Jeffrey P. Kharoufeh
Air Force Instititute of Technology

Follow this and additional works at: https://scholar.afit.edu/facpub

Part of the Operations Research, Systems Engineering and Industrial Engineering Commons

Recommended Citation

Sherman, N. P., & Kharoufeh, J. P. (2006). An M/M/1 retrial queue with unreliable server. Operations Research Letters, 34(6), 697–705. https://doi.org/10.1016/j.orl.2005.11.003

This Article is brought to you for free and open access by AFIT Scholar. It has been accepted for inclusion in Faculty Publications by an authorized administrator of AFIT Scholar. For more information, please contact richard.mansfield@afit.edu.

An M/M/1 Retrial Queue with Unreliable Server[1]

Nathan P. Sherman[2] and Jeffrey P. Kharoufeh[3]
Department of Operational Sciences
Air Force Institute of Technology

Abstract

We analyze an unreliable M/M/1 retrial queue with infinite-capacity orbit and normal queue. Retrial customers do not rejoin the normal queue but repeatedly attempt to access the server at i.i.d. intervals until it is found functioning and idle. We provide stability conditions as well as several stochastic decomposability results.

Keywords: Retrial queue, breakdowns, stochastic decomposition.

Final version appears in *Operations Research Letters*, 34 (6), 697-705.

[1]The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

[2]Voice: (937)-255-3636 x6589; Fax: (937)-656-4943; Email: Nathan.Sherman@afit.edu.

[3]Corresponding author. Air Force Institute of Technology, 2950 Hobson Way (AFIT/ENS), Wright Patterson AFB, OH 45433-7765, USA. Voice: (937)-255-3636 x4603; Fax: (937)-656-4943; Email: Jeffrey.Kharoufeh@afit.edu.

Report Documentation Page						Form Approved OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**
4. TITLE AND SUBTITLE **An M/M/1 Retrial Queue with Unreliable Server**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Institute of Technology (AFIT/ENS),Department of Operational Sciences,2950 Hobson Way,Wright Patterson AFB,OH,45433-7765**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We analyze an unreliable M/M/1 retrial queue with in?nite-capacity orbit and normal queue. Retrial customers do not rejoin the normal queue but repeatedly attempt to access the server at i.i.d. intervals until it is found functioning and idle. We provide stability conditions as well as several stochastic decomposability results.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**13**	

1 Introduction

In this paper, we study an M/M/1 retrial queue with an unreliable server whose orbit and normal queue have infinite storage capacity. Customers join the retrial orbit if and only if they are interrupted by a server breakdown. Retrial customers do not rejoin the normal queue, but rather attempt to access the server directly at random intervals independently of arrivals or other retrial customers. However, these interrupted customers can regain access to the server only when it is operational and idle and repeat service until they have been successfully processed.

Our primary motivation for studying this model stems from its interesting mathematical properties and its distinction from other commonly analyzed retrial queues. In particular, we are able to analyze the interaction between the orbit and the normal queue, an interaction that does not exist in the vast majority of retrial models that do not include an infinite- or nonzero-capacity normal queue. Under the assumed dynamics, we show that the steady-state orbit size and system size distributions possess a stochastic decomposition property. In particular, these random variables may be expressed as the sum of two independent random variables, one which corresponds to the same measure for an instantaneous feedback model (i.e., one with an infinite retrial rate) and the second which is a generalized negative binomial random variable. Furthermore, it will be shown that an interesting stability result emerges, namely that the normal queue may remain stable even if the stability condition for the entire system (and, in particular, for the orbit) is violated. Considered here are two types of breakdowns: *active breakdowns* which occur during a service cycle, and *idle breakdowns* which occur while the server is non-failed and idle. The times between customer arrivals, service completions, breakdowns, retrials, and repairs are assumed to be exponentially distributed random variables.

Queueing systems with breakdowns have been studied extensively in the literature as have retrial queues. However, the literature devoted to systems with both retrial queues and unreliable servers is comparatively sparse. The seminal papers in this area are [1] and [11]. All models considering retrial queues subject to server breakdowns assume an M/G/1/1 loss system with the exception of [6] which additionally considered a finite-capacity M/M/v/k queue for which the author derives sufficient ergodicity conditions but does not provide analytical distributions for queue lengths or other measures. Although [8] considered an M/G/1 retrial queue with infinite-capacity orbit and normal queue, the authors did not consider a server that is subject to breakdowns. In their model, arriving customers who find the server busy may join either the orbit or the normal queue.

For retrial models with no waiting room and server breakdowns, customers arriving to find the server unavailable (busy or failed) join the retrial orbit. Some models (cf. [2],[3], [18], [13], [14], [7], [16], [12]), force these customers into the orbit while others ([11], [6], [5], [4], [9], [17]) provide the option of joining the orbit or departing the system. With the exception of two cases ([3] and [16]), these models also either force, or provide the option for, in-service customers interrupted by a server failure to join the orbit. Our model is distinct in that arriving customers who find the server busy or failed join the normal queue whereas interrupted customers always join the orbit. A variety of failure types are considered in the literature including starting failures ([18], [13], [12]), vacations ([7], [14]), active breakdowns ([6], [16], [17]), and like our model, both active and idle breakdowns ([11], [2] [3], [5], [4], [9]). Most retrial orbits are assumed to behave as infinite-server queues with identical exponential service times; however some models (cf. [7], [12], and [17]) consider orbits as FCFS queues.

In this paper, we provide the steady-state joint distribution of the orbit size and normal queue size when the server is *idle* (operational and non-occupied), *failed* (non-operational and being repaired), or *busy* (operational and occupied). From these results, we obtain the joint probability generating function (p.g.f.) of the orbit size and normal queue size, and the p.g.f. of the overall system size (the total number of customers in orbit, normal queue and in service), independent of the server's status. We provide a necessary and sufficient condition for stability of the orbit and system and a necessary condition for stability of the normal queue. Moreover, we demonstrate the stochastic decomposability of the orbit and system size distributions and provide simple and intuitive expressions for the limiting distribution of the server's status and standard queueing performance measures. To our knowledge, this paper is the first to present such results for a retrial queue with breakdowns having an infinite-capacity orbit and normal queue.

The remainder of the paper is organized as follows. Section 2 provides the mathematical model description and discusses conditions for stability. In section 3, we provide the steady-state equations and the limiting distribution of server status, and also demonstrate the stochastic decomposition property for the orbit and system size distributions. Standard queueing performance measures are presented in section 4, while section 5 provides illustrative examples.

2 Model Description and Stability

Customers arrive to the system according to a Poisson process with rate $\lambda > 0$, and service times are i.i.d. exponential random variables with rate $\mu > 0$. Server failures occur at a constant rate $\xi > 0$, and the constant rate of repair is $\alpha > 0$. A customer interrupted by a server failure enters the retrial orbit and spends an exponential amount of time there with rate $\theta > 0$, after which it either enters service (if possible) or remains in the orbit for an additional exponentially distributed time with rate θ. Denote by Q_t the number of customers in the normal queue at time t, and let R_t denote the number of customers in the orbit at time t. The random variable X_t is the occupation status of the server given by

$$X_t = \begin{cases} 1, & \text{if the server is busy at time } t \\ 0, & \text{if the server is non-busy at time } t \end{cases}$$

while S_t describes the operational status of the server at time t defined by

$$S_t = \begin{cases} 1, & \text{if the server is non-failed at time } t \\ 0, & \text{if the server is failed at time } t \end{cases}.$$

The continuous-time stochastic process, $\{(Q_t, X_t, R_t, S_t) : t \geq 0\}$ describes the state of the system at time t. Let N_t denote the total number of customers in the system at time t (i.e., in orbit, normal queue, and in service). The process $\{N_t : t \geq 0\}$ describes the evolution of the system size over time. Denote by $\{\nu_n : n \geq 0\}$ the Markov chain embedded at the transition epochs of $\{N_t : t \geq 0\}$. We assume that as $t \to \infty$, $R_t \Rightarrow R$, $Q_t \Rightarrow Q$, and $N_t \Rightarrow N$, where $(\Rightarrow)$ denotes convergence in distribution.

The proportion of time the server is operational is $\alpha/(\alpha + \xi)$; thus, the effective service rate is $\alpha\mu/(\alpha + \xi)$ and $\lambda(\alpha + \xi)/\alpha\mu < 1$ is a necessary condition for system stability. As we shall see, this condition emerges naturally from Corollary 1 of section 3 and Corollary 3 of section 4. To show that $\lambda(\alpha+\xi)/\alpha\mu < 1$ is also sufficient, we need to prove that $\{\nu_n : n \geq 0\}$ is ergodic when $\lambda(\alpha+\xi)/\alpha\mu < 1$. It is easy to verify that $\{\nu_n : n \geq 0\}$ is irreducible and aperiodic; thus it remains to prove that it is positive recurrent. Pakes [15] proved that an irreducible and aperiodic Markov chain $\{\nu_n : n \geq 0\}$ is positive recurrent if $|\gamma_k| < \infty$ for all k and $\limsup_{k\to\infty} \gamma_k < 0$, where $\gamma_k \equiv E(\nu_{n+1} - \nu_n | \nu_n = k)$. In our model, $\gamma_0 = 1$ and $\gamma_k = (\lambda(\alpha+\xi) - \alpha\mu)/(\lambda(\alpha+\xi) + \alpha\mu)$, for all $k \geq 1$. Clearly, if $\lambda(\alpha+\xi)/\alpha\mu < 1$, then $|\gamma_k| < \infty$ for all k and $\limsup_{k\to\infty} \gamma_k < 0$. It will be shown in section 4 that this condition is more restrictive than the necessary condition for stability of the normal queue.

Define $\pi_{k,i,j,l}$ as the limiting probability that the system is in state (k, i, j, l), that is

$$\pi_{k,i,j,l} \equiv \lim_{t\to\infty} P(Q_t = k, X_t = i, R_t = j, S_t = l)$$

where the indices k, i, j, and l correspond to the normal queue size, the occupation status of the server (0 or 1), the orbit size, and the operational status of the server (0 or 1), respectively. The transform variables z_1 and z_2 correspond to the orbit size and the normal queue size, respectively. Let

$$\phi_{k,i,l}(z_1) \equiv \sum_{j=0}^{\infty} z_1^j \pi_{k,i,j,l}$$

denote the generating function of $\pi_{k,i,j,l}$ with respect to the orbit size and let

$$\psi_{i,l}(z_1, z_2) \equiv \sum_{k=0}^{\infty} z_2^k \phi_{k,i,l}(z_1)$$

denote the generating function of $\phi_{k,i,l}(z_1)$ with respect to the normal queue size. When the server is idle, failed, or busy, we respectively denote these p.g.f.s by $\phi_{0,0,1}(z_1)$, $\psi_{0,0}(z_1, z_2)$, and $\psi_{1,1}(z_1, z_2)$. The function $\phi'_{0,0,1}(z_1)$ is the first derivative of $\phi_{0,0,1}(z_1)$ with respect to z_1. Define $p(\cdot, \cdot)$ as the joint probability mass function (p.m.f.) of R and Q and let $q(\cdot)$ denote the p.m.f. of N. Then

$$G(z_1, z_2) \equiv \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} p(j, k) z_1^j z_2^k = \phi_{0,0,1}(z_1) + \psi_{0,0}(z_1, z_2) + \psi_{1,1}(z_1, z_2)$$

is the joint p.g.f. of the orbit and normal queue size, and

$$H(z) \equiv \sum_{j=0}^{\infty} q(j) z^j = \phi_{0,0,1}(z) + \psi_{0,0}(z, z) + z\psi_{1,1}(z, z)$$

denotes the p.g.f. of the system size. In section 3, we derive these transforms which provide the limiting distribution of server status and reveal some interesting and useful stochastic decomposition properties.

3 Steady-State Equations

In this section, we derive the steady-state joint distribution of the orbit and normal queue size when the server is idle, failed, or busy, respectively. Subsequently, we demonstrate the stochastic decomposability of these distributions independent of the server's status.

Theorem 1 *The generating functions, $\phi_{0,0,1}(z_1)$, $\psi_{0,0}(z_1, z_2)$, and $\psi_{1,1}(z_1, z_2)$, are given by*

$$\phi_{0,0,1}(z_1) = \frac{\alpha\mu - \lambda(\alpha+\xi)}{\mu(\alpha+\xi)} P(z_1)^c, \tag{1}$$

$$\begin{aligned} \psi_{0,0}(z_1, z_2) &= P(z_1)^{c+1} \\ &\times \frac{-\xi\{\lambda\xi(\alpha+\xi)z_1(z_1-z_2) + [\alpha\mu - \lambda(\alpha+\xi z_1)]\{(z_1-z_2)[\xi+\lambda(1-z_2)] + \mu(1-z_2)\}\}}{\mu(\alpha+\xi)\{[\xi z_2 - (\mu-\lambda z_2)(1-z_2)][\alpha+\lambda(1-z_2)] - \alpha\xi z_1\}}, \end{aligned} \tag{2}$$

and

$$\begin{aligned} \psi_{1,1}(z_1, z_2) &= P(z_1)^{c+1} \\ &\times \frac{-\lambda\{(1-z_2)[\alpha+\xi+\lambda(1-z_2)][\alpha\mu - \lambda(\alpha+\xi z_1)] + (z_1-z_2)\xi(\alpha+\xi)[\alpha+\lambda(1-z_2)]\}}{\mu(\alpha+\xi)\{[\xi z_2 - (\mu-\lambda z_2)(1-z_2)][\alpha+\lambda(1-z_2)] - \alpha\xi z_1\}}, \end{aligned} \tag{3}$$

where $P(z_1) = [\alpha\mu - \lambda(\alpha+\xi)]/[\alpha\mu - \lambda(\alpha+\xi z_1)]$ and $c = (\alpha+\xi)/\theta$.

Proof. Using the balance equations of this Markovian system, one can easily obtain

$$(\alpha+\lambda)\phi_{0,0,0}(z_1) = \xi\phi_{0,0,1}(z_1) + \xi z_1\phi_{0,1,1}(z_1) \tag{4}$$

$$(\alpha+\lambda)\phi_{k,0,0}(z_1) = \lambda\phi_{k-1,0,0}(z_1) + \xi z_1\phi_{k,1,1}(z_1) \tag{5}$$

$$(\lambda+\xi)\phi_{0,0,1}(z_1) + \theta z_1\phi'_{0,0,1}(z_1) = \alpha\phi_{0,0,0}(z_1) + \mu\phi_{0,1,1}(z_1) \tag{6}$$

$$(\lambda+\mu+\xi)\phi_{0,1,1}(z_1) = \alpha\phi_{1,0,0}(z_1) + \lambda\phi_{0,0,1}(z_1) + \mu\phi_{1,1,1}(z_1) + \theta\phi'_{0,0,1}(z_1) \tag{7}$$

$$(\lambda+\mu+\xi)\phi_{k,1,1}(z_1) = \alpha\phi_{k+1,0,0}(z_1) + \lambda\phi_{k-1,1,1}(z_1) + \mu\phi_{k+1,1,1}(z_1). \tag{8}$$

Multiplying both sides of Equation (4) by z_2^k for $k = 0$ and Equation (5) by z_2^k for $k \geq 1$ and summing over all $k \geq 0$, and similarly for Equations (7) and (8), we obtain, respectively, the following two equations:

$$[\alpha + \lambda(1 - z_2)]\psi_{0,0}(z_1, z_2) = \xi\phi_{0,0,1}(z_1) + \xi z_1\psi_{1,1}(z_1, z_2) \tag{9}$$

and

$$[\xi z_2 - (\mu - \lambda z_2)(1 - z_2)]\,\psi_{1,1}(z_1, z_2) + [\alpha\phi_{0,0,0}(z_1) + \mu\phi_{0,1,1}(z_1)] = \lambda z_2\phi_{0,0,1}(z_1) + \alpha\psi_{0,0}(z_1, z_2) + \theta z_2\phi'_{0,0,1}(z_1). \tag{10}$$

Using a technique employed in [6], we obtain another balance equation by equating the flow in and out of the set $E_n = \{(k, i, j, l) : k \leq n - 1\}$, $n \geq 1$ which leads to

$$\mu\phi_{n,1,1}(z_1) + \alpha\phi_{n,0,0}(z_1) = \lambda\phi_{n-1,1,1}(z_1) + \lambda\phi_{n-1,0,0}(z_1), \quad n \geq 1. \tag{11}$$

Multiplying both sides of Equation (11) by z_2^0 for $n = 0$ and by z_2^n for $n \geq 1$, summing over all n, and simplifying we obtain

$$(\mu - \lambda z_2)\psi_{1,1}(z_1, z_2) + (\alpha - \lambda z_2)\psi_{0,0}(z_1, z_2) = \alpha\phi_{0,0,0}(z_1) + \mu\phi_{0,1,1}(z_1). \tag{12}$$

We first obtain an expression for $\phi_{0,0,1}(z_1)$ by setting $z_2 = 1$ in equations (6), (9), (10), and (12) and solving

$$\phi'_{0,0,1}(z_1) = \frac{\lambda\xi(\alpha + \xi)}{\theta\,(\alpha\mu - \lambda(\alpha + \xi z_1))}\phi_{0,0,1}(z_1). \tag{13}$$

The general solution to this ordinary differential equation is

$$\phi_{0,0,1}(z_1) = \frac{C}{(\alpha\mu - \lambda(\alpha + \xi z_1))^{\frac{\alpha+\xi}{\theta}}}, \tag{14}$$

where C is a constant. Now substituting (14) into equations (9), (12), and (13), we obtain (1), (2) and (3) up to the multiplicative constant C. The final results are obtained after the normalization, $\phi_{0,0,1}(1) + \psi_{0,0}(1, 1) + \psi_{1,1}(1, 1) = 1$, which leads to

$$C = (\mu(\alpha + \xi))^{-1}\,[\alpha\mu - \lambda(\alpha + \xi)]^{c+1}$$

where $c = (\alpha+\xi)/\theta$. ■

Let p_I, p_F, and p_B denote the limiting probability that the server is idle, failed, or busy, respectively. A direct consequence of Theorem 1 is as follows.

Corollary 1 *The limiting probabilities p_I, p_F, and p_B are given by*

$$p_I = \lim_{z_1\to 1} \phi_{0,0,1}(z_1) = \frac{\alpha}{\alpha+\xi} - \frac{\lambda}{\mu},$$

$$p_F = \lim_{\substack{z_1\to 1\\ z_2\to 1}} \psi_{0,0}(z_1,z_2) = \frac{\xi}{\alpha+\xi},$$

and

$$p_B = \lim_{\substack{z_1\to 1\\ z_2\to 1}} \psi_{1,1}(z_1,z_2) = \frac{\lambda}{\mu}.$$

Corollary 1 reveals that $\lambda(\alpha+\xi)/\alpha\mu < 1$ is necessary to ensure system stability as noted in section 2. Theorem 1 also provides the means by which to obtain the joint distribution of the orbit and normal queue size, as well as the distribution of the overall system size, independent of server status.

Corollary 2 *The probability generating functions $G(z_1, z_2)$ and $H(z)$ are given by*

$$\begin{aligned} G(z_1,z_2) = \Bigg\{ & \frac{-[\alpha\mu-\lambda(\alpha+\xi z_1)][\alpha+\xi+\lambda(1-z_2)]\{\xi(z_1-z_2)+(1-z_2)[\mu+\lambda(1-z_2)]\}}{\mu(\alpha+\xi)\{[\xi z_2-(\mu-\lambda z_2)(1-z_2)][\alpha+\lambda(1-z_2)]-\alpha\xi z_1\}} \\ & + \frac{-\lambda\xi(\alpha+\xi)(z_1-z_2)[\alpha+\xi z_1+\lambda(1-z_2)]}{\mu(\alpha+\xi)\{[\xi z_2-(\mu-\lambda z_2)(1-z_2)][\alpha+\lambda(1-z_2)]-\alpha\xi z_1\}}\Bigg\} P(z_1)^{c+1} \end{aligned} \quad (15)$$

and

$$H(z) = \frac{[\alpha\mu-\lambda(\alpha+\xi z)][\alpha+\xi+\lambda(1-z)]}{(\alpha+\xi)\{(\mu-\lambda z)[\alpha+\lambda(1-z)]-\lambda\xi z\}} P(z)^{c+1}. \quad (16)$$

Using standard methods, (15) and (16) can be used to obtain the mth moment ($m \geq 1$) of R, Q, and N, respectively, as well as their limiting distributions.

The stochastic decomposition property has been observed for the system size distribution of many M/G/1 models including those with vacations, retrial queues, and breakdowns (cf. [18], [4], [9], and [12]). The property implies that the random variable of interest (e.g., orbit or system size) may be expressed as the sum of two independent random variables. We observe that Equations (1), (2), (3), (15), and (16) depend on the retrial rate θ only through the constant c. Allowing $\theta \to \infty$ yields a model in which retrial customers instantaneously attempt to re-access the server. Consequently, this shows that the orbit (system) size is the sum of two independent random variables: one is the orbit (system) size in the instantaneous feedback model and the second is a generalized negative binomial random variable. To see this, note that each of the aforementioned expressions shares a common multiplicative factor in Theorem 1. The following two propositions describe the decomposability of the orbit and system size distributions.

Proposition 1 *The stochastic decomposition for the orbit size is given by*

$$G(z_1, 1) = \left(\frac{1-r}{1-rz_1}\right)\left(\frac{1-r}{1-rz_1}\right)^c, \tag{17}$$

where $r = \lambda\xi/[\alpha(\mu-\lambda)]$.

Equation (17) is easily verified by setting $z_2 = 1$ in (15) and simplifying. The left-most term on the right-hand side of (17) is the generating function for R in the instantaneous feedback model (i.e., when $\theta \to \infty$), and the right-most term, $P(z_1)^c$, is the generating function of a generalized negative binomial distribution ([10]) with parameters r and c. Similar behavior may be observed for the system size distribution in the steady state as noted in Proposition 2

Proposition 2 *The stochastic decomposition of the system size is given by*

$$H(z) = B(z)\left(\frac{1-r}{1-rz_1}\right)^c, \tag{18}$$

where

$$B(z) \equiv \frac{[\alpha\mu - \lambda(\alpha+\xi z)][\alpha+\xi+\lambda(1-z)]}{(\alpha+\xi)\left\{(\mu-\lambda z)[\alpha+\lambda(1-z)] - \lambda\xi z\right\}}.$$

The quantity $B(z)$ in (18) corresponds to the generating function for N (the overall system size) in the instantaneous feedback model. In section 4, we provide standard queueing performance measures for this system.

4 Performance Measures

In this section, we use (15) and (16) to obtain the standard queueing performance measures noting that the mean system size and sojourn time may be decomposed into three components corresponding to the server, orbit, and normal queue measures.

Corollary 3 *The limiting mean orbit size, normal queue size, and system size are respectively given by*

$$E(R) = \frac{\alpha\lambda\xi[\mu(\mu+\xi-\lambda)+\lambda(\alpha+\xi)]}{\mu[\alpha\mu-\lambda(\alpha+\xi)][\alpha(\mu+\xi)-\lambda(\alpha+\xi)]} + \frac{\lambda\xi(\alpha+\xi)}{\theta[\alpha\mu-\lambda(\alpha+\xi)]}, \tag{19}$$

$$E(Q) = \lambda\frac{\mu\xi(\mu+\xi)+\lambda(\alpha+\xi)^2}{\mu(\alpha+\xi)[\alpha(\mu+\xi)-\lambda(\alpha+\xi)]}, \tag{20}$$

and

$$E(N) = \frac{\lambda[\mu\xi+(\alpha+\xi)^2]}{(\alpha+\xi)[\alpha\mu-\lambda(\alpha+\xi)]} + \frac{\lambda\xi(\alpha+\xi)}{\theta[\alpha\mu-\lambda(\alpha+\xi)]}. \tag{21}$$

Proof. The mean orbit size is obtained by evaluating $G'(z_1,1)$ at $z_1=1$, whereas the mean normal queue size is obtained by evaluating $G'(1,z_2)$ at $z_2=1$. In a similar manner, we obtain the mean system size by (16). ■

As $\xi \to 0$ in (19) through (21), we observe that $E(R) \to 0$, $E(Q) \to \lambda^2/\mu(\mu-\lambda)$, and $E(N) \to \lambda/(\mu-\lambda)$. These limiting values are consistent with results for the standard M/M/1 queue. Moreover, for λ, μ, α, and ξ fixed, the mean orbit size is bounded below by

$$\hat{E}(R) \equiv \lim_{\theta\to\infty} E(R) = \frac{\alpha\lambda\xi[\mu(\mu+\xi-\lambda)+\lambda(\alpha+\xi)]}{\mu[\alpha\mu-\lambda(\alpha+\xi)][\alpha(\mu+\xi)-\lambda(\alpha+\xi)]}.$$

It is worth noting that, if retrial customers are permitted to rejoin the normal queue, and we let $\theta\to\infty$, the model converges to the standard M/M/1 queue in which case the orbit is always empty and the orbit size distribution does not admit a stochastic decomposition.

By inspection of Equations (19) and (21) we observe that $\lambda(\alpha+\xi) < \alpha\mu$ is necessary for the stability of R (and N), and by (20) we see that $\lambda(\alpha+\xi) < \alpha(\mu+\xi)$ is necessary for the stability of Q. That is, the normal queue can be stable even if the orbit size stability condition is violated. Owing to the nature of the orbit dynamics, retrial customers are subordinate to normal customers and may be served only when the server is idle and operational. Hence, normal queue customers experience a greater effective service rate than do retrial customers, and thus, it is possible that the orbit may continue to grow while the normal queue remains stable.

We further note that the steady-state mean system size $E(N)$ may be conveniently decomposed into its constituent elements, namely the number of customers in orbit, normal queue, and service. The mean time spent in orbit, in the normal queue, and in the system are obtained by respectively dividing (19), (20), and (21) by λ.

5 Numerical Illustrations

We now illustrate the behavior of the mean orbit and normal queue size as functions of the traffic intensity, failure rate, repair rate, and retrial rate. In Figures 1 and 2, we respectively plot $E(R)$ and $E(Q)$, against traffic intensity λ/μ for four values of ξ when $\mu = 10$, $\alpha = 2$, and $\theta = 5$. In case $\xi = 0$, the model is equivalent to a standard M/M/1 queue with no retrials or failures. For $\xi = 0.5$, $\xi = 1.0$, and $\xi = 1.5$, $E(R)$ rapidly increases to the saturation point of $\alpha/(\alpha+\xi)$. Similarly in Figure 2, we observe that, for all choices of ξ, $E(Q)$ rapidly increases to the saturation point at which $\alpha(\mu+\xi) = \lambda(\alpha+\xi)$ (see Equation (20)). It is obvious that the orbit becomes unstable more quickly than the normal queue for the given failure and repair parameters.

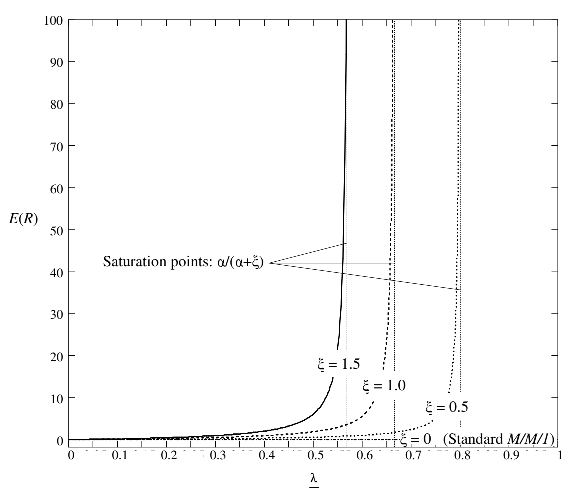

Figure 1: Mean orbit size ($\mu = 10$, $\alpha = 2$, $\theta = 5$).

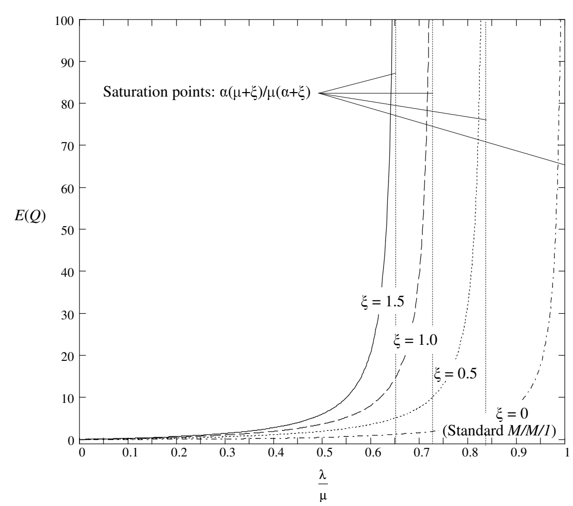

Figure 2: Mean normal queue size ($\mu = 10$, $\alpha = 2$, $\theta = 5$).

Acknowledgements. The authors acknowledge with gratitude the insightful comments provided by an anonymous referee.

References

[1] Aissani, A. (1988). On the M/G/1/1 queueing system with repeated orders and unreliable server. *Journal of Technology,* **6**, 98-123 (in French).

[2] Aissani, A. (1993). Unreliable queuing with repeated orders. *Microelectronics and Reliability,* **33** (14), 2093-2106.

[3] Aissani, A. (1994). A retrial queue with redundancy and unreliable server. *Queueing Systems,* **17** (3-4), 431-449.

[4] Aissani, A. and J.R. Artalejo (1998). On the single server retrial queue subject to breakdowns. *Queueing Systems,* **30** (3-4), 309-321.

[5] Anisimov, V. V. and K. L. Atadzhanov (1994). Diffusion approximation of systems with repeated calls and an unreliable server. *Journal of Mathematical Sciences,* **72** (2), 3032-3034.

[6] Artalejo, J. R. (1994). New results in retrial queueing systems with breakdown of the servers. *Statistica Neerlandica,* **48** (1), 23-36.

[7] Artalejo, J. R. (1997). Analysis of an M/G/1 queue with constant repeated attempts and server vacations. *Computers & Operations Research,* **24** (6), 493-504.

[8] Choi, B. D. and K. K. Park (1990). The M/G/1 retrial queue with Bernoulli schedule. *Queueing Systems,* **7** (2), 219-228.

[9] Djellab, N. V. (2002). On the M/G/1 retrial queue subjected to breakdowns. *RAIRO Operations Research,* **36** (4), 299-310.

[10] Jain, G. C. and P. C. Consul (1971). A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics,* **21** (4), 501-513.

[11] Kulkarni, V.G. and B. D. Choi (1990). Retrial queues with server subject to breakdowns and repairs. *Queueing Systems,* **7** (2), 191-208.

[12] Kumar, B. K., S. P. Madheswari, and A. Vijayakumar (2002). The M/G/1 retrial queue with feedback and starting failures. *Applied Mathematical Modelling,* **26** (11), 1057-1075.

[13] Li, H. and T. Yang (1995). A single-server retrial queue with server vacations and a finite number of input sources. *European Journal of Operations Research,* **85** (1), 149-160.

[14] Moutzoukis, E. and C. Langaris (1996). Non-preemptive priorities and vacations in a multiclass retrial queueing system. *Communications in Statistics-Stochastic Models,* **12** (3), 455-472.

[15] Pakes, A.G. (1969). Some conditions for ergodicity and recurrence of Markov chains. *Operations Research,* **17**, 1058-1061.

[16] Wang, J., Cao J., and Q. Li (2001). Reliability analysis of the retrial queue with server breakdowns and repairs. *Queueing Systems,* **38** (4), 363-380.

[17] Wu, X., P. Brill, M. Hlynka, and J. Wang (2005). An M/G/1 retrial queue with balking and retrials during service. *International Journal of Operational Research,* **1** (1/2) 30-51.

[18] Yang, T. and H. Li (1994). The M/G/1 retrial queue with the server subject to starting failures. *Queueing Systems,* **16** (1-2), 83-96.