Stochastic Analysis of a Single Server Retrial Queue with General Retrial Times

A. Gómez-Corral
Department of Statistics and Operations Research I, Faculty of Mathematics, Universidad Complutense de Madrid, 28040 Madrid, Spain

Received April 1998; revised November 1998; accepted 20 January 1999

Abstract: Retrial queueing systems are widely used in teletraffic theory and computer and communication networks. Although there has been a rapid growth in the literature on retrial queueing systems, the research on retrial queues with nonexponential retrial times is very limited. This paper is concerned with the analytical treatment of an $M/G/1$ retrial queue with general retrial times. Our queueing model is different from most single server retrial queueing models in several respectives. First, customers who find the server busy are queued in the orbit in accordance with an *FCFS* (first-come-first-served) discipline and only the customer at the head of the queue is allowed for access to the server. Besides, a retrial time begins (if applicable) only when the server completes a service rather upon a service attempt failure. We carry out an extensive analysis of the queue, including a necessary and sufficient condition for the system to be stable, the steady state distribution of the server state and the orbit length, the waiting time distribution, the busy period, and other related quantities. Finally, we study the joint distribution of the server state and the orbit length in non-stationary regime. © 1999 John Wiley & Sons, Inc. Naval Research Logistics 46: 561–581, 1999

1. INTRODUCTION

This paper deals with a single server retrial queue in which the retrial time is governed by a nonexponential distribution. Retrial queueing systems (queues with repeated attempts or queues with returning customers) permit no waiting in the normal sense. Instead a customer who finds all servers busy upon arrival is obliged to leave the service area and to come back to the system after a random amount of time. Between trials a customer is said to be in "orbit." These queueing systems are appropriate for communication networks, where a customer getting a busy signal tries his luck again at a later time.

Most papers on retrial queues assume that each customer seeks service independently of other customers in orbit after a random time exponentially distributed with rate $\mu > 0$. Thus, the probability of a repeated attempt during the interval $(t, t + \Delta t)$, given that j customers were in orbit at time t, is $j\mu\Delta t + o(\Delta t)$ as $\Delta t \rightarrow 0$. Yang and Templeton [21], Falin [8], and Kulkarni and Liang [15] provide extensive surveys of the results in this area. Nevertheless, there are other types of queueing situations in which the intervals separating successive repeated attempts are independent of the number of customers in the orbit, so that the probability of a

Contract grant sponsor: DGICYT; contract grant number: PB95-0416
Contract grant sponsor: European Commission; contract grant number: INTAS 96-0828

 CCC 0894-069X/99/050561-21

repeated attempt during $(t, t + \Delta t)$, given that the orbit is not empty at time t, is $\mu\Delta t + o(\Delta t)$ as $\Delta t \to 0$. This second type of retrial discipline was introduced by Fayolle [10], who investigated a telephone exchange model as an $M/M/1$ retrial queue in which the orbiting customer who finds the server busy joins the tail of the queue. In Choi, Park, and Pearce [4], the authors considered an $M/M/1$ queue where the retrial time has a general distribution and only the customer at the head of the queue is allowed to retry for service. The variant with general service times and exponentially distributed intervals separating the retrial epochs was dealt with extensively by Martin and Artalejo [16]. For a detailed review of the main results and the literature on the two above retrial disciplines, the reader is referred to Falin and Templeton [9].

Until now, retrial queueing systems with general service times and nonexponential retrial time distribution have received little attention. The first work on the $M/G/1$ retrial queue with general retrial times is due to Kapyrin [12], who assumed that each customer in orbit generates a stream of repeated attempts that is independent of the rest of the customers in orbit and the server state. However, his methodology was found to be incorrect (see Falin [7]). Later, a method convenient for practical applications, based on the stochastic decomposition property, was proposed in Yang et al. [22]. In that paper, the authors developed an approximation method for the calculation of the steady state performance measures of the queue described by Kapyrin [12]. Besides, Pourbabai has published several papers (see Pourbabai [18] and its references) in which an alternative approximation of retrial queues with general retrial times based on the corresponding loss models is considered.

Our objective here is to develop an exhaustive analysis of an $M/G/1$ retrial queue with nonexponential retrial times. The main characteristic of this queue is that, at any service completion, a competition between an exponential law and a general retrial time distribution determines the next customer who accesses the service facility. Thus, the retrial discipline does not depend on the orbit length. Our model is related to many service systems where, after the end of the service time, the processor shall expend a random time interval to locate the next item to be processed. As a related work, Neuts and Ramalhoto [17] analyzed a queueing model in which, at the end of the transmission time, the server is required to search for customers.

The paper is organized as follows. Section 2 gives the mathematical description and the necessary and sufficient condition for the system to be stable. The steady state distribution of the server state and the orbit length is studied in Section 3. The waiting time distribution is investigated in Section 4. The busy period, the number of customers served and other related quantities are analyzed in Section 5. Section 6 deals with the analysis of the transient distribution of the server state and the orbit length.

2. MODEL DESCRIPTION AND ERGODICITY CONDITION

We consider a single server queue in which customers arrive according to a Poisson stream of rate $\lambda > 0$. If, upon arrival, the server is idle, then the service of the arriving customer commences immediately. Otherwise, the customer leaves the service area and enters a group of blocked customers called "orbit" in accordance with an *FCFS* discipline. We shall assume that only the customer at the head of the orbit queue is allowed for access to the server. When a service is completed, the access from the orbit to the server is governed by an arbitrary law with common probability distribution function $A(x)$ $[A(0) = 0]$, of density function $a(x)$ and Laplace–Stieltjes transform $\alpha(\theta)$. The service times are independent with common probability distribution function $B(x)$ $[B(0) = 0]$, of density function $b(x)$, Laplace–Stieltjes transform $\beta(\theta)$, and first moments $\beta_k = (-1)^k\beta^k(0)$, $k = 1, 2, 3$. Interarrival times, retrial times, and service times are assumed to be mutually independent.

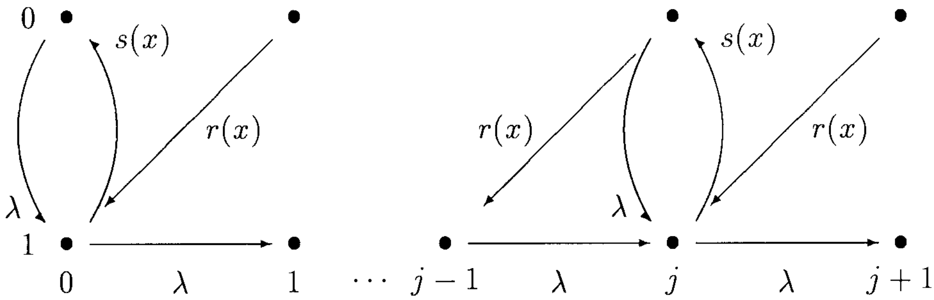

Figure 1. State-transition-rate diagram.

From this description, it is clear that the evolution of the retrial queue exhibits an alternating sequence of idle and busy periods of the server. At any service completion, the server becomes free and a competition between an exponential law of rate λ and the general retrial time distribution determines the next customer who accesses the service facility.

The state of the system at time t can be described by the Markov process $\{X(t),\ t \geq 0\} = \{(C(t),\ Q(t),\ \xi_0(t),\ \xi_1(t)),\ t \geq 0\}$, where $C(t)$ denotes the server state (0 or 1 according to the server is free or busy, respectively) and $Q(t)$ is the number of customers in orbit at time t. If $C(t) = 0$ and $Q(t) > 0$, then $\xi_0(t)$ represents the elapsed retrial time. If $C(t) = 1$, then $\xi_1(t)$ corresponds to the elapsed time of the customer being served. The transitions among states are illustrated in Figure 1. The functions $r(x)$ and $s(x)$ are the conditional completion rates for repeated attempt and for service at time x, respectively; i.e., $r(x) = a(x)(1 - A(x))^{-1}$ and $s(x) = b(x)(1 - B(x))^{-1}$.

We first study the necessary and sufficient condition for the system to be stable. To see this, in the following theorem we establish the ergodicity of the embedded Markov chain at departure epochs.

THEOREM 1: Let Q_n be the orbit length at the time of the nth departure, $n \geq 1$. Then, $\{Q_n,\ n \geq 1\}$ is ergodic if and only if $\lambda\beta_1 < \alpha(\lambda)$.

PROOF: To prove ergodicity, we shall use the following Foster's criterion: An irreducible and aperiodic Markov chain is ergodic if there exists a nonnegative function $f(j)$, $j \in \mathbb{N}$, and $\varepsilon > 0$ such that the mean drift $\chi_j = E[f(Q_{n+1}) - f(Q_n)/Q_n = j]$ is finite for all $j \in \mathbb{N}$ and $\chi_j \leq -\varepsilon$ for all $j \in \mathbb{N}$, except perhaps a finite number of j.

In our case, we consider the function $f(j) = j$. Then, we obtain that $\chi_j = \lambda\beta_1 - (1 - \delta_{0j})\alpha(\lambda)$ for $j \geq 0$, where δ_{ab} denotes Kronecker's delta. Clearly, the inequality $\lambda\beta_1 < \alpha(\lambda)$ is a sufficient condition for ergodicity.

The same inequality is also necessary for ergodicity. As noted in Sennot, Humblet, and Tweedie [19, Theorem 1], we can guarantee nonergodicity if the Markov chain $\{Q_n,\ n \geq 1\}$ satisfies Kaplan's condition, $\chi_j < +\infty$ for all $j \geq 0$ and there exists $j_0 \in \mathbb{N}$ such that $\chi_j \geq 0$ for $j \geq j_0$. Notice that, in our case, Kaplan's condition is fulfilled because $r_{ij} = 0$ for $j < i - 1$ and $i > 0$, where $R = (r_{ij})$ is the one-step transition matrix of $\{Q_n,\ n \geq 1\}$. Then, $\lambda\beta_1 \geq \alpha(\lambda)$ implies the nonergodicity of the Markov chain. □

Since the arrival stream is a Poisson process, it can be shown from Burke's theorem (see Cooper [6, p. 187]) that the steady state probabilities of $\{(C(t),\ Q(t)),\ t \geq 0\}$ exist and are positive if and only if $\lambda\beta_1 < \alpha(\lambda)$.

From the mean drift $\chi_j = \lambda\beta_1 - \alpha(\lambda)$ for $j \geq 1$, we have the reasonable conclusion that the expected number of ordinary customers who arrive per service interval equals $\lambda\beta_1$ and the expected number of orbiting customers who enter service at the epoch at which a service starts, given that the previous service time left j customers in orbit, equals $\alpha(\lambda)$. For stability we require $\lambda\beta_1 < \alpha(\lambda)$, and so Theorem 1 indicates that ordinary customers must arrive per service interval more slowly than orbiting customers can enter service at the epoch at which a service starts (on the average).

3. STEADY STATE DISTRIBUTION OF THE SERVER STATE AND THE ORBIT LENGTH

In this section we study the steady state distribution of the system. For the process $\{X(t), t \geq 0\}$, we define the probability $P_0(t) = P\{C(t) = 0, Q(t) = 0\}$ and the probability densities $P_j(t, x)\,dx = P\{C(t) = 0, Q(t) = j, x \leq \xi_0(t) < x + dx\}$ for $t \geq 0$, $x \geq 0$ and $j \geq 1$, and $Q_j(t, x)\,dx = P\{C(t) = 1, Q(t) = j, x \leq \xi_1(t) < x + dx\}$ for $t \geq 0$, $x \geq 0$ and $j \geq 0$. Following the routine procedure of the method of supplementary variables (see, for instance, Keilson, Cozzolino, and Young [13]), we obtain the equations that govern the dynamics of the system:

$$\frac{dP_0(t)}{dt} + \lambda P_0(t) = \int_0^\infty Q_0(t, x)s(x)\,dx, \tag{1}$$

$$\frac{\partial P_j(t, x)}{\partial t} + \frac{\partial P_j(t, x)}{\partial x} + (\lambda + r(x))P_j(t, x) = 0, \qquad j \geq 1, \tag{2}$$

$$\frac{\partial Q_j(t, x)}{\partial t} + \frac{\partial Q_j(t, x)}{\partial x} + (\lambda + s(x))Q_j(t, x) = (1 - \delta_{0j})\lambda Q_{j-1}(t, x), \qquad j \geq 0, \tag{3}$$

$$P_j(t, 0) = \int_0^\infty Q_j(t, x)s(x)\,dx, \qquad j \geq 1, \tag{4}$$

$$Q_j(t, 0) = \delta_{0j}\lambda P_0(t) + (1 - \delta_{0j})\lambda \int_0^\infty P_j(t, x)\,dx + \int_0^\infty P_{j+1}(t, x)r(x)\,dx, \qquad j \geq 0, \tag{5}$$

and the normalizing condition

$$P_0(t) + \sum_{j=1}^\infty \int_0^\infty P_j(t, x)\,dx + \sum_{j=0}^\infty \int_0^\infty Q_j(t, x)\,dx = 1. \tag{6}$$

We assume that the condition $\lambda\beta_1 < \alpha(\lambda)$ is fulfilled and set $P_0 = \lim_{t\to\infty} P_0(t)$, $P_j(x) = \lim_{t\to\infty} P_j(t, x)$ for $x \geq 0$ and $j \geq 1$, and $Q_j(x) = \lim_{t\to\infty} Q_j(t, x)$ for $x \geq 0$ and $j \geq 0$. Hence, letting $t \to \infty$ in Eqs. (1)–(6), we obtain

$$\lambda P_0 = \int_0^\infty Q_0(x)s(x)\,dx, \tag{7}$$

$$\frac{dP_j(x)}{dx} + (\lambda + r(x))P_j(x) = 0, \qquad j \geq 1, \tag{8}$$

$$\frac{dQ_j(x)}{dx} + (\lambda + s(x))Q_j(x) = (1 - \delta_{0j})\lambda Q_{j-1}(x), \qquad j \geq 0, \tag{9}$$

$$P_j(0) = \int_0^\infty Q_j(x)s(x)\,dx, \qquad j \geq 1, \tag{10}$$

$$Q_j(0) = \delta_{0j}\lambda P_0 + (1 - \delta_{0j})\lambda \int_0^\infty P_j(x)\,dx + \int_0^\infty P_{j+1}(x)r(x)\,dx, \qquad j \geq 0, \tag{11}$$

$$P_0 + \sum_{j=1}^\infty \int_0^\infty P_j(x)\,dx + \sum_{j=0}^\infty \int_0^\infty Q_j(x)\,dx = 1. \tag{12}$$

To solve Eqs. (7)–(12), we define the generating functions $\phi_0(z, x) = \sum_{j=1}^\infty P_j(x)z^j$ and $\phi_1(z, x) = \sum_{j=0}^\infty Q_j(x)z^j$.

The following theorem fully describes the steady state distribution of the system.

THEOREM 2: If $\lambda\beta_1 < \alpha(\lambda)$, then the steady state distribution of $\{X(t), t \geq 0\}$ is given by

$$P_0 = 1 - \frac{\lambda\beta_1}{\alpha(\lambda)}, \tag{13}$$

$$\phi_0(z, x) = \frac{\lambda(\alpha(\lambda) - \lambda\beta_1)z(1 - \beta(\lambda - \lambda z))}{\alpha(\lambda)(\alpha(\lambda)(1 - z)\beta(\lambda - \lambda z) - z(1 - \beta(\lambda - \lambda z)))}(1 - A(x))e^{-\lambda x}, \tag{14}$$

$$\phi_1(z, x) = \frac{\lambda(\alpha(\lambda) - \lambda\beta_1)(1 - z)}{\alpha(\lambda)(1 - z)\beta(\lambda - \lambda z) - z(1 - \beta(\lambda - \lambda z))}(1 - B(x))e^{-(\lambda-\lambda z)x}. \tag{15}$$

The generating function $K(z) = P_0 + \int_0^\infty \phi_0(z, x)\,dx + z\int_0^\infty \phi_1(z, x)\,dx$ of the number of customers in the system is given by

$$K(z) = \frac{(\alpha(\lambda) - \lambda\beta_1)(1 - z)\beta(\lambda - \lambda z)}{\alpha(\lambda)(1 - z)\beta(\lambda - \lambda z) - z(1 - \beta(\lambda - \lambda z))}. \tag{16}$$

PROOF: Multiplying Eqs. (8)–(11) by z^j and summing over j, we obtain the following equations:

$$\frac{\partial \phi_0(z, x)}{\partial x} + (\lambda + r(x))\phi_0(z, x) = 0, \tag{17}$$

$$\frac{\partial \phi_1(z, x)}{\partial x} + (\lambda - \lambda z + s(x))\phi_1(z, x) = 0, \tag{18}$$

$$\phi_0(z, 0) + \int_0^\infty \phi_1(0, x)s(x)\, dx = \int_0^\infty \phi_1(z, x)s(x)\, dx, \tag{19}$$

$$z\phi_1(z, 0) = \lambda z\left(P_0 + \int_0^\infty \phi_0(z, x)\, dx \right) + \int_0^\infty \phi_0(z, x)r(x)\, dx, \tag{20}$$

$$P_0 + \int_0^\infty \phi_0(1, x)\, dx + \int_0^\infty \phi_1(1, x)\, dx = 1. \tag{21}$$

Solving the partial differential equations (17) and (18), we have that

$$\phi_0(z, x) = \phi_0(z, 0)(1 - A(x))e^{-\lambda x}, \qquad x \geq 0, \tag{22}$$

$$\phi_1(z, x) = \phi_1(z, 0)(1 - B(x))e^{-(\lambda - \lambda z)x}, \qquad x \geq 0. \tag{23}$$

Combining (7) and (19)–(23), we obtain after algebraic manipulation (13)–(15).

Finally, the expression (16) for $K(z)$ is easily obtained by integrating Eqs. (14) and (15) from $x = 0$ to $x = \infty$. □

REMARK 1: The generating function $K(z)$ of the number of customers in the system at an arbitrary time point coincides with the generating function $\Pi(z)$ of the embedded Markov chain at departure epochs, which can be obtained from the following Kolmogorov equations for the stationary probabilities $\pi_j = P\{Q_n = j\}$, $j \geq 0$:

$$\pi_j = \pi_0 k_j + (1 - \delta_{0j})(1 - \alpha(\lambda)) \sum_{n=1}^{j} \pi_n k_{j-n} + \alpha(\lambda) \sum_{n=1}^{j+1} \pi_n k_{j+1-n}, \qquad j \geq 0,$$

where $k_j = \int_0^\infty e^{-\lambda x}(\lambda x)^j (j!)^{-1}\, dB(x)$ for $j \geq 0$. This result can also be obtained directly by noting that any stochastic process whose sample functions are, almost all, step functions with unit jumps has the same steady state distribution just prior to its points of increase as it does just after its points of decrease, when this steady state distribution exists (see Cooper [6, p. 187]). Thus, we have that $\pi_j = \delta_{0j}P_0 + (1 - \delta_{0j})(\int_0^\infty P_j(x)\, dx + \int_0^\infty Q_{j-1}(x)\, dx)$ for $j \geq 0$.

REMARK 2: Our retrial queue can be thought of as an $M/G/1$ *queue with generalized vacations* (see Fuhrmann and Cooper [11]) in which the vacation begins at the end of each service time and the server is turned on again according to a competition between an exponential law of rate $\lambda > 0$ (corresponding to primary customers) and the general retrial time distribution. As noted in Artalejo and Gómez-Corral [3], the steady state number of customers present in the system at a random point in time is distributed as the sum of two independent random variables: the steady state number of customers present in the standard $M/G/1$ queue and the steady state number of customers present in the queue under study given that the server is not available. This property is known in the literature as the $M/G/1$ decomposition property or the stochastic decomposition law. We thus get that

$$K(z) = \tilde{K}(z) \frac{P_0 + \phi_0(z)}{\lim_{t\to\infty} P\{C(t) = 0\}},$$

where $\phi_0(z) = \int_0^\infty \phi_0(z, x)\, dx$ and $\tilde{K}(z)$ is the generating function on the number of customers present in the standard $M/G/1$ queue; i.e., $\tilde{K}(z) = (1 - \lambda\beta_1)(1 - z)\beta(\lambda - \lambda z)(\beta(\lambda - \lambda z) - z)^{-1}$, for $|z| \le 1$.

We summarize in the following corollary some performance measures related to the steady state distribution of the system.

COROLLARY 1: Let M_k^0 and M_k^1 denote the factorial moments defined as $M_k^0 = \sum_{j=k}^\infty \prod_{n=0}^{k-1} (j - n) \int_0^\infty P_j(x)\, dx$ and $M_k^1 = \sum_{j=k}^\infty \prod_{n=0}^{k-1} (j - n) \int_0^\infty Q_j(x)\, dx$ for $k \ge 1$. If $\lambda\beta_1 < \alpha(\lambda)$, then

$$\lim_{t\to\infty} P\{C(t) = 0\} = 1 - \lambda\beta_1, \tag{24}$$

$$M_1^0 = \frac{\lambda(1 - \alpha(\lambda))(\lambda\beta_2 + 2(1 - \lambda\beta_1)\beta_1)}{2(\alpha(\lambda) - \lambda\beta_1)}, \tag{25}$$

$$M_1^1 = \frac{\lambda^2(\alpha(\lambda)\beta_2 + 2(1 - \alpha(\lambda))\beta_1^2)}{2(\alpha(\lambda) - \lambda\beta_1)}, \tag{26}$$

$$\begin{aligned} M_2^0 = \frac{\lambda^2(1 - \alpha(\lambda))}{6(\alpha(\lambda) - \lambda\beta_1)^2} &(2\lambda(\alpha(\lambda) - \lambda\beta_1)\beta_3 + 3(\lambda^2\beta_2 + 2(\alpha(\lambda)(1 - \lambda\beta_1) \\ &+ \lambda(1 - \alpha(\lambda))\beta_1))\beta_2 + 12(1 - \alpha(\lambda))(1 - \lambda\beta_1)\beta_1^2), \end{aligned} \tag{27}$$

$$\begin{aligned} M_2^1 = \frac{\lambda^3}{6(\alpha(\lambda) - \lambda\beta_1)^2} &(2\alpha(\lambda)(\alpha(\lambda) - \lambda\beta_1)\beta_3 + 3\alpha(\lambda)(\lambda\beta_2 \\ &+ 4(1 - \alpha(\lambda))\beta_1)\beta_2 + 12(1 - \alpha(\lambda))^2\beta_1^3). \end{aligned} \tag{28}$$

The proof follows by routine differentiation of the generating functions $\phi_0(z)$ and $\phi_1(z) = \int_0^\infty \phi_1(z, x)\, dx$, and thus we omit the details.

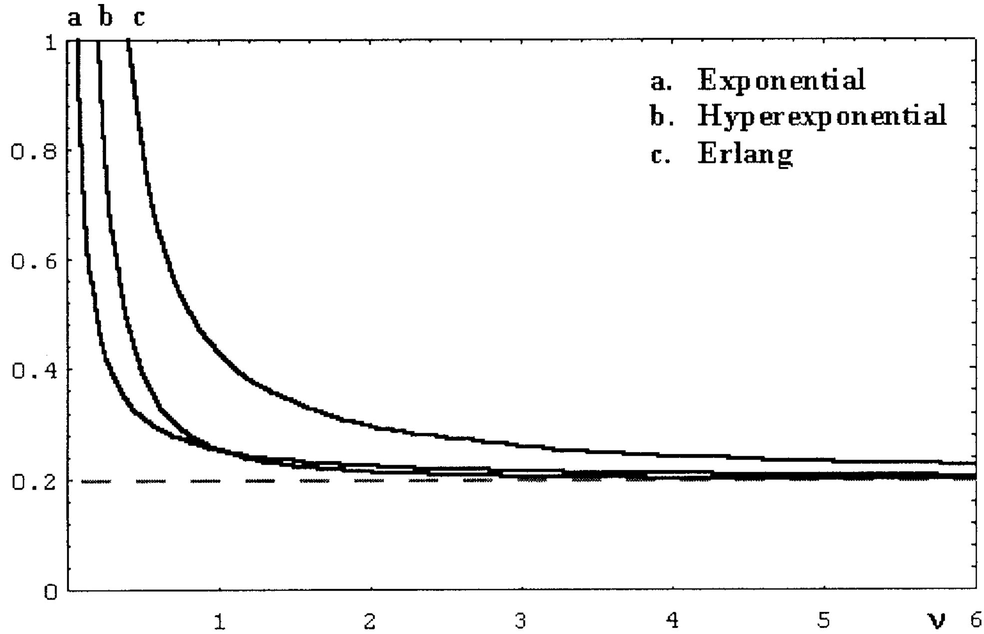

Figure 2. ρ versus the parameter ν.

We now discuss a basis parameter ρ, which is commonly referred to as the *utilization factor*. In the case of the standard $M/G/1$ queue, the utilization factor ρ_0 is the ratio of the rate at which "work" enters the system to the maximum rate (capacity) at which the system can perform this work; i.e., ρ_0 may be interpreted as $\rho_0 = E[\text{fraction of busy servers}] = \lambda\beta_1$ (see Kleinrock [14, p. 18]). The interpretation here is that $\rho = 1 - P_0 = \lambda\beta_1/\alpha(\lambda)$ is the fraction of time during which the server is busy and/or there are customers in orbit. Besides, ρ_0 agrees with the limiting probability that the server is busy in our retrial queue [see Eq. (24)].

The effect of the retrial time distribution on the utilization factor and the mean number of customers in the system in steady state (i.e., $M_1^0 + M_1^1 + \lim_{t\to\infty} P\{C(t) = 1\} = \lambda(\lambda\beta_2 + 2(1 - \lambda\beta_1)\beta_1)(2(\alpha(\lambda) - \lambda\beta_1))^{-1}$) is illustrated in Figures 2 and 3. We consider retrial queues with arrival rate λ equals 0.3, where service times follow the gamma(α, β) distribution with $\alpha = 1.0$ and $\beta = 0.65$ (i.e., $\beta_1 = 0.65$ and $\beta_2 = 1.0725$). It is assumed that retrial times follow the exponential distribution $[a(x) = \nu e^{-\nu x},\ x > 0]$, Erlang distribution $[a(x) = \nu^k e^{-\nu x} x^{k-1}/(k - 1)!,\ x > 0$, where $k = 3]$, and hyperexponential distribution $[a(x) = p\nu e^{-\nu x} + (1 - p)\nu^2 e^{-\nu^2 x},\ x > 0$, where $p = 0.25]$. Then, ρ and the mean number of customers in the system in steady state are plotted versus the parameter ν. The curves show that ρ and $M_1^0 + M_1^1 + \lim_{t\to\infty} P\{C(t) = 1\}$ converge to the corresponding quantities related to the standard $M/G/1$ queue, as $\nu \to \infty$, as we expected when the mean retrial time converges to zero. Besides, they show that $\rho_0 \le \rho$ and the mean number of customers in the standard $M/G/1$ queue is a lower bound for the mean number of customers in our retrial queue. Substantially the same remark could be made for the expected waiting time and the expected length of a busy period.

4. WAITING TIME PROCESS

Let $W(t)$ be the waiting time of a primary customer who arrives at the system at time t. The waiting time is the sojourn time that a customer spends waiting in the orbit. The waiting time

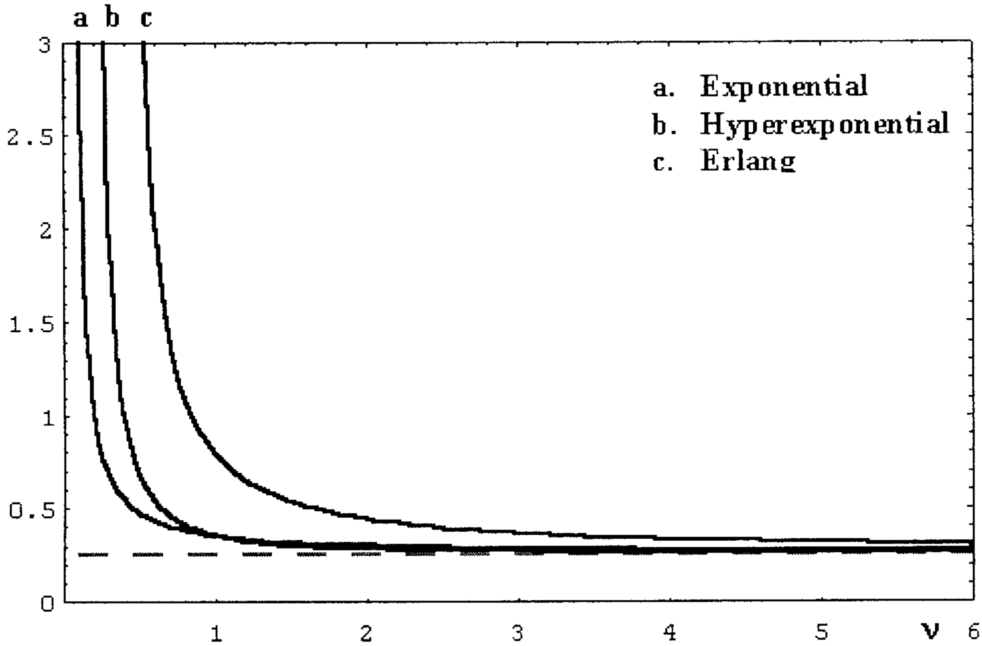

Figure 3. Mean number of customers in the system in steady state versus the parameter ν.

process depends on the orbit discipline for accessing to the server, so we shall assume that only the customer at the head of the orbit queue is allowed for access to the server.

The main characteristics of the waiting time process are:

$W_i(t)$, the total amount of time in $[t, t + W(t))$ during which the server is in state i, for $i \in \{0, 1\}$,

$N(t)$, the number of customers served during $[t, t + W(t))$.

In this section, we are interested in studying the joint distribution of the random vector $(W_0(t), W_1(t), N(t))$, in steady state, and its corresponding first moment. To that end, we define the joint transform $\Phi(\theta_0, \theta_1, z)$ as follows:

$$\Phi(\theta_0, \theta_1, z) = E[\exp\{-\theta_0 W_0(t) - \theta_1 W_1(t)\} z^{N(t)}], \tag{29}$$

for $\mathrm{Re}(\theta_0) \geq 0$, $\mathrm{Re}(\theta_1) \geq 0$, $|z| \leq 1$.

THEOREM 3: If $\lambda\beta_1 < \alpha(\lambda)$, then the joint transform of $(W_0(t), W_1(t), N(t))$ under *FCFS* orbit discipline is given by

$$\Phi(\theta_0, \theta_1, z) = 1 - \lambda\beta_1 + \lambda(\alpha(\lambda) - \lambda\beta_1) \frac{\tilde{\omega}(\theta_0, \theta_1, z)(1 - \tilde{\omega}(\theta_0, \theta_1, z))\zeta(\theta_0, \theta_1, z)}{(\theta_1 - \lambda + \lambda\tilde{\omega}(\theta_0, \theta_1, z))\beta(\theta_1)\gamma(\theta_0, \theta_1, z)}, \tag{30}$$

where

$$\zeta(\theta_0, \theta_1, z) = \beta(\lambda - \lambda\tilde{\omega}(\theta_0, \theta_1, z)) - \beta(\theta_1),$$

$$\tilde{\omega}(\theta_0, \theta_1, z) = \frac{(\theta_0 + \lambda)\alpha(\theta_0 + \lambda)\beta(\theta_1)z}{\theta_0 + \lambda - \lambda(1 - \alpha(\theta_0 + \lambda))\beta(\theta_1)z},$$

$$\gamma(\theta_0, \theta_1, z) = \alpha(\lambda)(1 - \tilde{\omega}(\theta_0, \theta_1, z))\beta(\lambda - \lambda\tilde{\omega}(\theta_0, \theta_1, z))$$
$$-\tilde{\omega}(\theta_0, \theta_1, z)(1 - \beta(\lambda - \lambda\tilde{\omega}(\theta_0, \theta_1, z))).$$

PROOF: The formula of the total probability allows us to express $\Phi(\theta_0, \theta_1, z)$ as

$$\Phi(\theta_0, \theta_1, z) = 1 - \lambda\beta_1 + \int_0^\infty \sum_{j=0}^\infty E[e^{-\theta_0 W_0(t) - \theta_1 W_1(t)} z^{N(t)} / (C(t-), Q(t-), \xi_1(t-)) = (1, j, u)] \times d_u P\{C(t-) = 1, Q(t-) = j, \xi_1(t-) < u\}. \quad (31)$$

We now define $\Phi_{1ju}(\theta_0, \theta_1, z) = E[\exp\{-\theta_0 W_0(t) - \theta_1 W_1(t)\} z^{N(t)} / (C(t-), Q(t-), \xi_1(t-)) = (1, j, u)]$ for $\mathrm{Re}(\theta_0) \geq 0$, $\mathrm{Re}(\theta_1) \geq 0$, $|z| \leq 1$, $j \geq 0$, and $u \geq 0$. To study $\Phi_{1ju}(\theta_0, \theta_1, z)$, we first observe that the conditional distribution of the residual time of the customer in service at time t, $\eta(t)$, is given by

$$E[e^{-\theta_1 \eta(t)} / (C(t-), Q(t-), \xi_1(t-)) = (1, j, u)] = \frac{e^{\theta_1 u}}{1 - B(u)} \int_u^\infty e^{-\theta_1 x}\, dB(x). \quad (32)$$

Hence, $\Phi_{1ju}(\theta_0, \theta_1, z)$ can be expressed as

$$\Phi_{1ju}(\theta_0, \theta_1, z) = \frac{z e^{\theta_1 u}}{1 - B(u)} \int_u^\infty e^{-\theta_1 x}\, dB(x)\, E[\exp\{-\theta_0 \tau_{j+1}^0 - \theta_1 \tau_{j+1}^1\} z^{N_{j+1}}], \quad (33)$$

for $j \geq 0$ and $u \geq 0$, where τ_{j+1}^i is the total amount of time in $[t + \eta(t), t + W(t))$ during which the server is in state i, given that $Q(t + \eta(t)) = j + 1$, for $i \in \{0, 1\}$, and N_{j+1} is the total number of customers served during the interval $(t + \eta(t), t + W(t))$, given that $Q(t + \eta(t)) = j + 1$.

Since *FCFS* discipline is assumed in the orbit, we obtain

$$E[\exp\{-\theta_0 \tau_{j+1}^0 - \theta_1 \tau_{j+1}^1\} z^{N_{j+1}}] = E[\exp\{-\theta_0 \tau_1^0 - \theta_1 \tau_1^1\} z^{N_1}] \times (z\beta(\theta_1) E[\exp\{-\theta_0 \tau_1^0 - \theta_1 \tau_1^1\} z^{N_1}])^j, \qquad j \geq 0. \quad (34)$$

The competition between the exponential law of rate $\lambda > 0$ and the general retrial time distribution is analyzed by conditioning on the number of primary customers that get service before the customer at the head of the orbit. Therefore, it is readily seen that

$$E[\exp\{-\theta_0\tau_1^0 - \theta_1\tau_1^1\}z^{N_1}] = \frac{(\theta_0 + \lambda)\alpha(\theta_0 + \lambda)}{\theta_0 + \lambda - \lambda(1 - \alpha(\theta_0 + \lambda))\beta(\theta_1)z}. \tag{35}$$

Inserting (33)–(35) into (31), we find that

$$\Phi(\theta_0, \theta_1, z) = 1 - \lambda\beta_1 + \frac{\tilde{\omega}(\theta_0, \theta_1, z)}{\beta(\theta_1)} \int_0^\infty \frac{1}{1 - B(u)} \int_u^\infty e^{-\theta_1(x-u)}\, dB(x)\phi_1(\tilde{\omega}(\theta_0, \theta_1, z), u)\, du.$$

Taking into account of (13)–(15), we deduce (30). □

With the help of $\Phi(\theta_0, \theta_1, z)$ we can get various performance measures of the waiting time process.

COROLLARY 2: If $\lambda\beta_1 < \alpha(\lambda)$, then the expected value of the random vector $(W_0(t), W_1(t), N(t))$ is given by

$$E[W_0(t)] = \frac{(1 - \alpha(\lambda))(\lambda\beta_2 + 2(1 - \lambda\beta_1)\beta_1)}{2(\alpha(\lambda) - \lambda\beta_1)},$$

$$E[W_1(t)] = \frac{\lambda(\alpha(\lambda)\beta_2 + 2(1 - \alpha(\lambda))\beta_1^2)}{2(\alpha(\lambda) - \lambda\beta_1)},$$

$$E[N(t)] = \frac{\lambda(\lambda\beta_2 + 2(1 - \lambda\beta_1)\beta_1)}{2(\alpha(\lambda) - \lambda\beta_1)}.$$

Note that $E[N(t)]$ coincides with the mean number of customers in the system in steady state given in Section 3. Moreover, it can be seen that Little's formula $M_1^0 + M_1^1 + \lim_{t\to\infty} P\{C(t) = 1\} = \lambda E[\tilde{W}(t)]$ holds, where $\tilde{W}(t)$ denotes the time spent in the system; i.e., $E[\tilde{W}(t)] = \beta_1 + E[W_0(t)] + E[W_1(t)]$, $M_1^0 = \lambda E[W_0(t)]$, and $M_1^1 = \lambda E[W_1(t)]$.

5. BUSY PERIODS AND IDLE TIMES

A system busy period is defined as the period that starts at an epoch when an arriving customer finds an empty system and ends at the next departure epoch at which the system is empty. The main characteristics of a system busy period are:

L, the length of a system busy period,
I, the number of service completions occurring during $(0, L]$.

The system busy period of a retrial queue consists of alternating service periods and periods in which the server is idle and there are customers in orbit. Nevertheless, the system busy period can also be expressed in terms of a sequence of orbit busy periods and periods of competition between the service time distribution and the Poisson input process (see Artalejo and Falin [2]).

First, we will investigate two characteristics of the orbit, namely, the orbit idle period and the orbit busy period. These characteristics are defined as follows:

i. *The orbit idle period* is the interval of time from the epoch of beginning of service of a customer who is alone in the orbit, makes a new attempt to get service, and finds the server idle, until the epoch at which a primary customer finds the server busy and is obliged to join the orbit.

ii. *The orbit busy period* is the interval of time from the epoch when a primary customer arrives and finds the server busy and the orbit idle, until the next epoch at which a repeated attempt finds the server idle and the orbit becomes empty.

The following random variables are related to the orbit idle period and the orbit busy period:

L_i, the length of an orbit idle period,
I_i, the number of service completions occurring during $(0, L_i]$,
L_b, the length of an orbit busy period,
I_b, the number of service completions occurring during $(0, L_b]$.

Let us assume that (L_b, I_b) are the characteristics of an orbit busy period that is preceded by an orbit idle period with characteristics (L_i, I_i). For the random vector (L_i, I_i, L_b, I_b) we define the transform:

$$\varphi(\theta, \omega, z, u) = E[\exp\{-\theta L_i - \omega L_b\} z^{I_i} u^{I_b}], \tag{36}$$

for $\mathrm{Re}(\theta) \geq 0$, $\mathrm{Re}(\omega) \geq 0$, $|z| \leq 1$, and $|u| \leq 1$. In order to obtain $\varphi(\theta, \omega, z, u)$, we need to consider the auxiliary random variable Y defined as the duration of the competition between a service time and the Poisson arrival stream which takes place just before the beginning of the orbit busy period. The joint distribution of (L_i, I_i, Y) does not depend on the process of repeated attempts. Thus, we have that (see Artalejo and Falin [2, Theorem 1]):

$$E[e^{-\theta L_i} z^{I_i} I_{\{Y=\tau\}}] = \frac{\lambda(\theta + \lambda)(1 - B(\tau))}{\theta + \lambda - \lambda\beta(\theta + \lambda)z} e^{-(\theta+\lambda)\tau}, \tag{37}$$

for $\mathrm{Re}(\theta) > 0$, $|z| \leq 1$ and $\tau \geq 0$.

The following lemma will be used later on. The proof follows from Rouche's theorem.

LEMMA 1: Define the function

$$F(\omega, u, \sigma) = u(\lambda\sigma + (\omega + \lambda - \lambda\sigma)\alpha(\omega + \lambda))\beta(\omega + \lambda - \lambda\sigma) - (\omega + \lambda)\sigma,$$

for $\mathrm{Re}(\omega) > 0$, $|u| \leq 1$, and $|\sigma| \leq 1$. Then for each fixed value of (ω, u) the function F has a unique zero $\sigma = \sigma(\omega, u)$ in the unit disk $|\sigma| \leq 1$. Moreover, if $\lambda\beta_1 < \alpha(\lambda)$, then $\sigma(0, 1) = 1$ and

$$\left(\frac{\partial\sigma(\omega,\,1)}{\partial\omega}\right)_{\omega=0} = \frac{1}{\lambda}\left(1 - \frac{1}{\alpha(\lambda) - \lambda\beta_1}\right) \qquad \text{and} \qquad \left(\frac{\partial\sigma(0,\,u)}{\partial u}\right)_{u=1} = \frac{1}{\alpha(\lambda) - \lambda\beta_1}.$$

The distribution of (L_i, I_i, L_b, I_b) is stated in the following theorem.

THEOREM 4: The joint transform of (L_i, I_i, L_b, I_b) is given by:

$$\varphi(\theta,\,\omega,\,z,\,u) = \frac{\lambda(\theta+\lambda)\sigma(\omega,\,u)(\beta(\theta+\lambda) - \beta(\omega+\lambda-\lambda\sigma(\omega,\,u)))}{(\theta+\lambda-\lambda\beta(\theta+\lambda)z)(\omega-\theta-\lambda\sigma(\omega,\,u))\beta(\omega+\lambda-\lambda\sigma(\omega,\,u))}, \tag{38}$$

where $\sigma(\omega, u)$ is the solution of the equation $(\omega + \lambda)\sigma = u(\lambda\sigma + (\omega + \lambda - \lambda\sigma)\alpha(\omega + \lambda))\beta(\omega + \lambda - \lambda\sigma)$ in the unit disk, which is continuous at the point $(\omega, u) = (0, 1)$.

PROOF: Conditioning on the value of Y and taking into account that (L_b, I_b) depends on (L_i, I_i) only through Y, we find that

$$\varphi(\theta,\,\omega,\,z,\,u) = \int_0^\infty E[e^{-\theta L_i} z^{I_i} I_{\{Y=\tau\}}]\pi_\tau(\omega,\,u)\,d\tau, \tag{39}$$

where $\pi_\tau(\omega, u) = E[e^{-\omega L_b} u^{I_b}/Y = \tau]$.

To determine $\pi_\tau(\omega, u)$ we denote by $I(t)$ the number of service completions in $(0, t]$ and define the probability densities $P^b_{jn}(t, x)\,dx = P\{C(t) = 0,\, Q(t) = j,\, x \le \xi_0(t) < x + dx,\, I(t) = n,\, L_b > t\}$ for $j \ge 1$ and $n \ge 1$, and $Q^b_{jn}(t, x)\,dx = P\{C(t) = 1,\, Q(t) = j,\, x \le \xi_1(t) < x + dx,\, I(t) = n,\, L_b > t\}$ for $j \ge 1$ and $n \ge 0$. Then, it should be noted that

$$\pi^\tau_n(t) = \frac{d}{dt}P\{I(t) = n,\, L_b \le t\} = \int_0^\infty P^b_{1n}(t,\,x)r(x)\,dx, \qquad n \ge 1. \tag{40}$$

With the help of the method of supplementary variables, we obtain the Kolmogorov equations for $P^b_{jn}(t, x)$ and $Q^b_{jn}(t, x)$:

$$\frac{\partial P^b_{jn}(t,\,x)}{\partial t} + \frac{\partial P^b_{jn}(t,\,x)}{\partial x} + (\lambda + r(x))P^b_{jn}(t,\,x) = 0, \qquad j \ge 1, \quad n \ge 1, \tag{41}$$

$$\frac{\partial Q^b_{jn}(t,\,x)}{\partial t} + \frac{\partial Q^b_{jn}(t,\,x)}{\partial x} + (\lambda + s(x))Q^b_{jn}(t,\,x) = (1 - \delta_{1j})\lambda Q^b_{j-1,n}(t,\,x), \qquad j \ge 1, \quad n \ge 0, \tag{42}$$

$$P^b_{jn}(t,\,0) = \int_0^\infty Q^b_{j,n-1}(t,\,x)s(x)\,dx, \qquad j \ge 1, \quad n \ge 1, \tag{43}$$

$$Q^b_{j0}(t, 0) = 0, \qquad j \geq 1, \tag{44}$$

$$Q^b_{jn}(t, 0) = \lambda \int_0^\infty P^b_{jn}(t, x)\, dx + \int_0^\infty P^b_{j+1,n}(t, x) r(x)\, dx, \qquad j \geq 1, \quad n \geq 1. \tag{45}$$

The initial conditions are

$$P^b_{jn}(0, x) = 0, \qquad j \geq 1, \quad n \geq 1, \qquad Q^b_{jn}(0, x) = \delta_{1j}\delta_{0n}\delta(x - \tau), \qquad j \geq 1, \quad n \geq 0, \tag{46}$$

where $\delta(x)$ is the Dirac delta.

To solve (40)–(46) we introduce the generating functions $\tilde{p}(\omega, x, u, v) = \sum_{n=1}^\infty u^n \sum_{j=1}^\infty v^j \tilde{P}_{jn}(\omega, x)$ and $\tilde{q}(\omega, x, u, v) = \sum_{n=0}^\infty u^n \sum_{j=1}^\infty v^j \tilde{Q}_{jn}(\omega, x)$, where $\tilde{P}_{jn}(\omega, x)$ and $\tilde{Q}_{jn}(\omega, x)$ are the Laplace–Stieltjes transforms of $P^b_{jn}(t, x)$ and $Q^b_{jn}(t, x)$, respectively.

From (40)–(46) we obtain after algebraic manipulation:

$$\frac{\partial \tilde{p}(\omega, x, u, v)}{\partial x} + (\omega + \lambda + r(x))\tilde{p}(\omega, x, u, v) = 0, \tag{47}$$

$$\frac{\partial \tilde{q}(\omega, x, u, v)}{\partial x} + (\omega + \lambda - \lambda v + s(x))\tilde{q}(\omega, x, u, v) = v\delta(x - \tau), \tag{48}$$

$$\tilde{p}(\omega, 0, u, v) = u \int_0^\infty \tilde{q}(\omega, x, u, v) s(x)\, dx, \tag{49}$$

$$v(\pi_\tau(\omega, u) + \tilde{q}(\omega, 0, u, v)) = \lambda v \int_0^\infty \tilde{p}(\omega, x, u, v)\, dx + \int_0^\infty \tilde{p}(\omega, x, u, v) r(x)\, dx. \tag{50}$$

To prove (50), we note that $\pi_\tau(\omega, u)$ may be expressed in terms of the Laplace–Stieltjes transforms $\tilde{\pi}^\tau_n(\omega) = \int_0^\infty e^{-\omega t}\pi^\tau_n(t)\, dt$, $n \geq 1$, as $\pi_\tau(\omega, u) = \sum_{n=1}^\infty \tilde{\pi}^\tau_n(\omega) u^n$.

From (47) and (48) we find that $\tilde{p}(\omega, x, u, v)$ and $\tilde{q}(\omega, x, u, v)$ depend upon x as follows:

$$\tilde{p}(\omega, x, u, v) = \tilde{p}(\omega, 0, u, v)(1 - A(x))e^{-(\omega+\lambda)x}, \tag{51}$$

$$\tilde{q}(\omega, x, u, v) = \left(\tilde{q}(\omega, 0, u, v) + \frac{ve^{(\omega+\lambda-\lambda v)\tau}}{1 - B(\tau)} I_{\{x \geq \tau\}}\right)(1 - B(x))e^{-(\omega+\lambda-\lambda v)x}. \tag{52}$$

Combining (47), (48), (51), and (52), and after rearrangement:

$$(u(\lambda v + (\omega + \lambda - \lambda v)\alpha(\omega + \lambda))\beta(\omega + \lambda - \lambda v) - (\omega + \lambda)v)\tilde{p}(\omega, 0, u, v)$$
$$= (\omega + \lambda)uv\left(\beta(\omega + \lambda - \lambda v)\pi_\tau(\omega, u) - \frac{v\beta_\tau(\omega + \lambda - \lambda v)}{1 - B(\tau)}\, e^{(\omega+\lambda-\lambda v)\tau}\right), \tag{53}$$

where $\beta_\tau(\omega) = \int_\tau^\infty e^{-\omega x}\, dB(x)$.

Replacing $v = \sigma(\omega, u)$ in (53), we get

$$\pi_\tau(\omega, u) = \frac{\sigma(\omega, u)\beta_\tau(\omega + \lambda - \lambda\sigma(\omega, u))}{\beta(\omega + \lambda - \lambda\sigma(\omega, u))(1 - B(\tau))}\, e^{(\omega+\lambda-\lambda\sigma(\omega,u))\tau}. \tag{54}$$

Then, we deduce from (37), (39) and (54) the expression (38). □

The definition of the orbit busy period allows to study in more detail the structure of the system busy period. Let N_b be the number of orbit busy periods which take place in $(0, L]$.

THEOREM 5: The joint transform of the random vector (L, I, N_b) is given by

$$\psi(\theta, u, v) = E[e^{-\theta L}u^I v^{N_b}] = \frac{u\beta(\theta + \lambda)\beta(\theta + \lambda - \lambda\sigma(\theta, u))}{v\beta(\theta + \lambda) + (1 - v)\beta(\theta + \lambda - \lambda\sigma(\theta, u))}, \tag{55}$$

where $\sigma(\theta, u)$ is the zero of $F(\theta, u, \sigma)$ in Lemma 1.

PROOF: Let $\tilde{Y}$ be the duration of the competition between the arrival of a new primary customer and a service time, given that this competition ends in completion of the service time. The Laplace–Stieltjes transform of $\tilde{Y}$ is

$$E[e^{-\theta\tilde{Y}}] = \frac{\beta(\theta + \lambda)}{\beta(\lambda)}. \tag{56}$$

Conditioning on the value of N_b and taking into account the independence among the random vectors of the sequence $\{(Y_n, L_b^n, I_b^n), 0 \le n \le N_b\}$, where Y_n, L_b^n, and I_b^n are distributed as Y, L_b, and I_b, respectively, we get

$$\psi(\theta, u, v) = \sum_{n=0}^{\infty} \beta(\lambda)(1 - \beta(\lambda))^n\left(v\int_0^\infty E[e^{-\theta(Y+L_b)}u^{I_b}/Y = \tau]f_Y(\tau)\, d\tau\right)^n uE[e^{-\theta\tilde{Y}}], \tag{57}$$

where $f_Y(\tau)$ is the density function of Y; i.e., $f_Y(\tau) = (1 - \beta(\lambda))^{-1}\lambda(1 - B(\tau))e^{-\lambda\tau}$, for $\tau > 0$.

Combining (54), (56), and (57), and rearranging gives (55). □

For the sake of completeness, we also compute some results in the following corollary.

COROLLARY 3: The joint transforms of (L_b, I_b) and (L, I), and the generating function of N_b are

$$E[e^{-\omega L_b}u^{I_b}] = \frac{\lambda\sigma(\omega, u)(\beta(\lambda) - \beta(\omega + \lambda - \lambda\sigma(\omega, u)))}{(1 - \beta(\lambda))(\omega - \lambda\sigma(\omega, u))\beta(\omega + \lambda - \lambda\sigma(\omega, u))},$$

$$E[e^{-\theta L}u^{I}] = u\beta(\theta + \lambda - \lambda\sigma(\theta, u)) \qquad \text{and} \qquad E[v^{N_b}] = \frac{\beta(\lambda)}{1 - (1 - \beta(\lambda))v}.$$

If $\lambda\beta_1 < \alpha(\lambda)$, then

$$E[L_b] = \frac{\lambda\beta_1 - (1 - \beta(\lambda))\alpha(\lambda)}{\lambda(1 - \beta(\lambda))(\alpha(\lambda) - \lambda\beta_1)}, \qquad E[I_b] = \frac{\beta(\lambda)\lambda\beta_1}{(1 - \beta(\lambda))(\alpha(\lambda) - \lambda\beta_1)},$$

$$E[L] = \frac{\beta_1}{\alpha(\lambda) - \lambda\beta_1}, \qquad E[I] = \frac{\alpha(\lambda)}{\alpha(\lambda) - \lambda\beta_1}, \qquad \text{and} \qquad E[N_b] = \frac{1 - \beta(\lambda)}{\beta(\lambda)}.$$

6. TRANSIENT DISTRIBUTION OF THE SERVER STATE AND THE ORBIT LENGTH

In this section we consider the transient distribution

$$P_{0n}(t) = P\{C(t) = 0, Q(t) = 0, I(t) = n\}, \qquad n \geq 0,$$

$$P_{jn}(t, x)\,dx = P\{C(t) = 0, Q(t) = j, x \leq \xi_0(t) < x + dx, I(t) = n\}, \qquad j \geq 1, \quad n \geq 0,$$
$$Q_{jn}(t, x)\,dx = P\{C(t) = 1, Q(t) = j, x \leq \xi_1(t) < x + dx, I(t) = n\}, \qquad j \geq 0, \quad n \geq 0,$$

of the process $\{(C(t), Q(t), \xi_0(t), \xi_1(t), I(t)), t \geq 0\}$ in nonstationary regime. For convenience, we will assume that at time $t = 0$ the system is empty and $I(0) = 0$.

We first introduce some notation. Let us define $\hat{P}_{jn}(t, x)\,dx = P\{C(t) = 0, Q(t) = j, x \leq \xi_0(t) < x + dx, I(t) = n, L > t\}$ for $j \geq 1$ and $n \geq 1$, and $\hat{Q}_{jn}(t, x)\,dx = P\{C(t) = 1, Q(t) = j, x \leq \xi_1(t) < x + dx, I(t) = n, L > t\}$ for $j \geq 0$ and $n \geq 0$. We also define the Laplace–Stieltjes transforms

$$\hat{p}(\theta, x, z, u) = \int_0^\infty e^{-\theta t}p(t, x, z, u)\,dt \qquad \text{and} \qquad \hat{q}(\theta, x, z, u) = \int_0^\infty e^{-\theta t}q(t, x, z, u)\,dt,$$

of the generating functions

$$p(t, x, z, u) = \sum_{n=1}^\infty u^n \sum_{j=1}^\infty z^j\hat{P}_{jn}(t, x) \qquad \text{and} \qquad q(t, x, z, u) = \sum_{n=0}^\infty u^n \sum_{j=0}^\infty z^j\hat{Q}_{jn}(t, x).$$

The Laplace–Stieltjes transforms $\hat{p}(\theta, x, z, u)$ and $\hat{q}(\theta, x, z, u)$ are determined by considering the differential difference equations given by Eqs. (1)–(5) with the absorbing barrier imposed at the zero system size and the initial condition $(C(0), Q(0), I(0)) = (1, 0, 0)$. Then, we get that

$$\pi_n(t) = \int_0^\infty Q_{0,n-1}(t, x)s(x)\, dx, \qquad n \geq 1, \tag{58}$$

where $\pi_n(t)$ is the density function of $\Pi_n(t) = P\{I(t) = n, L \leq t\}$ for $n \geq 1$. Thus, the necessary equations are

$$\frac{\partial \hat{P}_{jn}(t, x)}{\partial t} + \frac{\partial \hat{P}_{jn}(t, x)}{\partial x} + (\lambda + r(x))\hat{P}_{jn}(t, x) = 0, \qquad j \geq 1, \quad n \geq 1, \tag{59}$$

$$\frac{\partial \hat{Q}_{jn}(t, x)}{\partial t} + \frac{\partial \hat{Q}_{jn}(t, x)}{\partial x} + (\lambda + s(x))\hat{Q}_{jn}(t, x) = (1 - \delta_{0j})\lambda \hat{Q}_{j-1,n}(t, x), \qquad j \geq 0, \quad n \geq 0, \tag{60}$$

$$\hat{P}_{jn}(t, 0) = \int_0^\infty \hat{Q}_{j,n-1}(t, x)s(x)\, dx, \qquad j \geq 1, \quad n \geq 1, \tag{61}$$

$$\hat{Q}_{j0}(t, 0) = 0, \qquad j \geq 0, \tag{62}$$

$$\hat{Q}_{jn}(t, 0) = (1 - \delta_{0j})\lambda \int_0^\infty \hat{P}_{jn}(t, x)\, dx + \int_0^\infty \hat{P}_{j+1,n}(t, x)r(x)\, dx, \qquad j \geq 0, \quad n \geq 1, \tag{63}$$

with the initial conditions

$$\hat{P}_{jn}(0, x) = 0, \qquad j \geq 1, \quad n \geq 1, \qquad \hat{Q}_{jn}(0, x) = \delta_{0j}\delta_{0n}\delta(x), \qquad j \geq 0, \quad n \geq 0. \tag{64}$$

Verbatim repetition of the proofs of Theorems 2 and 4 of Sections 3 and 5, respectively, gives the following result.

THEOREM 6: The Laplace–Stieltjes transforms $\hat{p}(\theta, x, z, u)$ and $\hat{q}(\theta, x, z, u)$ are given by

$$\hat{p}(\theta, x, z, u) = (1 - A(x))e^{-(\theta+\lambda)x} \times \frac{(\theta + \lambda)zu(\beta(\theta + \lambda - \lambda z) - \beta(\theta + \lambda - \lambda\sigma(\theta, u)))}{(\theta + \lambda)z - u(\lambda z + (\theta + \lambda - \lambda z)\alpha(\theta + \lambda))\beta(\theta + \lambda - \lambda z)}, \tag{65}$$

$$\hat{q}(\theta, x, z, u) = (1 - B(x))e^{-(\theta+\lambda-\lambda z)x} \times \frac{(\theta + \lambda)z - u(\lambda z + (\theta + \lambda - \lambda z)\alpha(\theta + \lambda))\beta(\theta + \lambda - \lambda\sigma(\theta, u))}{(\theta + \lambda)z - u(\lambda z + (\theta + \lambda - \lambda z)\alpha(\theta + \lambda))\beta(\theta + \lambda - \lambda z)}, \quad (66)$$

where $\sigma(\theta, u)$ is the zero of $F(\theta, u, z)$ in Lemma 1.

It should be noted that the numerators and denominators of (65) and (66) have singularities at the point $z = \sigma(\theta, u)$ which can be determined by continuity.

Next, we introduce the Laplace–Stieltjes transforms $p_0^*(\theta, u)$, $p^*(\theta, x, z, u)$, and $q^*(\theta, x, z, u)$ of the generating functions $p_0(t, u) = \sum_{n=0}^{\infty} u^n P_{0n}(t)$, $p(t, x, z, u) = \sum_{j=1}^{\infty} z^j \sum_{n=0}^{\infty} u^n P_{jn}(t, x)$, and $q(t, x, z, u) = \sum_{j=0}^{\infty} z^j \sum_{n=0}^{\infty} u^n Q_{jn}(t, x)$, respectively.

To get expressions for $p_0^*(\theta, u)$, $p^*(\theta, x, z, u)$, and $q^*(\theta, x, z, u)$, we employ the method of collective marks (see Kleinrock [14]). Fix $\theta > 0$ and introduce a Poisson process of rate θ. The events of this process are called *catastrophes*. Besides, we introduce two processes to mark customers by painting each arriving customer (respectively, served customer) with *red* color (respectively, *blue* color) with probability z, $z \in [0, 1]$ (respectively, with probability y, $y \in [0, 1]$), and *green* color (respectively, *yellow* color) with probability $1 - z$ (respectively, $1 - y$). We assume that these auxiliary processes are mutually independent and are also independent of the functioning of the queue. Then, $\theta p_0^*(\theta, u)$ can be thought of as the probability of the event $\mathcal{A}$ defined as *at the time* τ_θ, *when the first catastrophe occurred, the system was empty and all customers served were marked in blue.* Similarly, we can think of $\theta p^*(\theta, x, z, u)\, dx$ (respectively, $\theta q^*(\theta, x, z, u)\, dx$) as the probability of the event $\mathcal{B}$ (respectively, $\mathcal{C}$) defined as *at time* τ_θ, *when the first catastrophe occurred, the server was free (respectively, busy), the elapsed retrial time* $\xi_0(\tau_\theta) \in [x, x + dx)$ (*respectively, the elapsed service time* $\xi_1(\tau_\theta) \in [x, x + dx)$), *all customers in the system were marked in red and all customers served were marked in blue.*

Conditioning on the number of busy periods occurring before τ_θ, χ_θ, we get that $\theta p_0^*(\theta, u) = \sum_{n=0}^{\infty} P\{\mathcal{A}, \chi_\theta = n\}$, $\theta p^*(\theta, x, z, u) = \sum_{n=0}^{\infty} P\{\mathcal{B}, \chi_\theta = n\}$ and $\theta q^*(\theta, x, z, u) = \sum_{n=0}^{\infty} P\{\mathcal{C}, \chi_\theta = n\}$. Now, the probabilities $P\{\mathcal{A}, \chi_\theta = n\}$, $P\{\mathcal{B}, \chi_\theta = n\}$ and $P\{\mathcal{C}, \chi_\theta = n\}$ can be written as

$$P\{\mathcal{A}, \chi_\theta = n\} = \frac{\theta}{\theta + \lambda}\left(\frac{\lambda}{\theta + \lambda} E[e^{-\theta L}u^I]\right)^n, \qquad n \geq 0,$$

$$P\{\mathcal{B}, \chi_\theta = n\} = \frac{\lambda\theta\hat{p}(\theta, x, z, u)}{\theta + \lambda}\left(\frac{\lambda}{\theta + \lambda} E[e^{-\theta L}u^I]\right)^n, \qquad n \geq 0,$$

$$P\{\mathcal{C}, \chi_\theta = n\} = \frac{\lambda\theta\hat{q}(\theta, x, z, u)}{\theta + \lambda}\left(\frac{\lambda}{\theta + \lambda} E[e^{-\theta L}u^I]\right)^n, \qquad n \geq 0.$$

This analysis yields the following result.

THEOREM 7: The Laplace–Stieltjes transforms $p_0^*(\theta, u)$, $p^*(\theta, x, z, u)$, and $q^*(\theta, x, z, u)$ of the transient distribution of the process $\{(C(t), Q(t), \xi_0(t), \xi_1(t), I(t)), t \geq 0\}$ in nonstationary regime are given by the following formulas:

$$p_0^*(\theta, u) = \frac{1}{\theta + \lambda - \lambda u \beta(\theta + \lambda - \lambda\sigma(\theta, u))},$$

$$p^*(\theta, x, z, u) = \frac{\lambda \hat{p}(\theta, x, z, u)}{\theta + \lambda - \lambda u \beta(\theta + \lambda - \lambda\sigma(\theta, u))},$$

$$q^*(\theta, x, z, u) = \frac{\lambda \hat{q}(\theta, x, z, u)}{\theta + \lambda - \lambda u \beta(\theta + \lambda - \lambda\sigma(\theta, u))},$$

where $\hat{p}(\theta, x, z, u)$ and $\hat{q}(\theta, x, z, u)$ were given in Theorem 6 and $\sigma(\theta, u)$ is the zero of $F(\theta, u, \sigma)$ in Lemma 1.

Now we note how to compute the time dependent distribution of the server state and the orbit length. In the interest of brevity, we give some simple description of a multivariate generalization of the Euler and lattice-Poisson algorithms studied by Abate and Whitt [1]. In particular, we focus on the transient functions $\{P_{0n}(t),\ t \geq 0,\ n \geq 0\}$, which we may calculate by numerically inverting the bidimensional transform $p_0^*(\theta, u)$ given in Theorem 7; i.e.,

$$p_0^*(\theta, u) = \int_0^\infty \sum_{n=0}^\infty P_{0n}(t) e^{-\theta t} u^n \, dt.$$

The bivariate mixed Poisson summation formula of a function $F(t, n)$ defined for $t \in \mathbb{R}$ and $n \in \mathbb{Z}$ is given by

$$\sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} F\left(t + \frac{2\pi j}{h}, n + km\right) = \frac{h}{2\pi m} \sum_{j=-\infty}^{+\infty} \sum_{k=-m/2}^{m/2-1} F^*\left(jh, \frac{2\pi k}{m}\right) \exp\left\{-i\left(jht + \frac{2\pi kn}{m}\right)\right\},$$

where $F^*(\theta, v)$ is the Fourier transform of $F(t, n)$. If we replace $F(t, n)$ by the function $P_{0n}(t)e^{-at}r^n$ when $t \geq 0$ and $n \geq 0$, where $a > 0$ and $r \in (0, 1)$, then $F^*(\theta, v) = p_0^*(a - i\theta, re^{iv})$. Assuming that $h = \pi(tl_1)^{-1}$, $m = 2l_2 n$, and $a = A(2tl_1)^{-1}$, the above equation becomes

$$P_{0n}(t) = g(t, n) - \varepsilon,$$

where the function $g(t, n)$ is given by

$$g(t, n) = \frac{1}{2l_1 t} \exp\left\{\frac{A}{2l_1}\right\} \sum_{j_1=1}^{l_1} \sum_{j=-\infty}^{+\infty} (-1)^j \exp\left\{-\frac{ij_1\pi}{l_1}\right\} \frac{1}{2l_2 n r_2^n}$$
$$\times \sum_{k_1=0}^{l_2-1} \sum_{k=-n}^{n-1} (-1)^k \exp\left\{-\frac{ik_1\pi}{l_2}\right\} p_0^*\left(\frac{A}{2l_1 t} - \frac{ij_1\pi}{l_1 t} - \frac{ij\pi}{t}, r \exp\left\{\frac{i\pi(k_1 + l_2 k)}{l_2 n}\right\}\right)$$

and the *aliasing* error ε may be bounded as

$$|\varepsilon| \leq \frac{e^{-A} + r^{2l_2n} - e^{-A}r^{2l_2n}}{(1 - e^{-A})(1 - r^{2l_2n})}.$$

Then, the roundoff error and ε may be controlled by the parameters (A, r, l_1, l_2), and we may efficiently calculate each infinite series of the form $\sum_{j=-\infty}^{+\infty} (-1)^j a_j$ from finitely many terms by exploiting the Euler transformation. For a detailed study of algorithms for numerically inverting multidimensional transforms of probability distributions of continuous random variables (Laplace–Stieltjes transforms) and/or discrete random variables (generating functions) the reader is referred to Choudhury, Luncantoni, and Whitt [5] and its references.

To conclude the paper, we point out that, although our derivations employ an approach based on the method of collective marks and the method of supplementary variables under the assumption of density function $a(x)$ and $A(0) = 0$, the results given in this paper are also consistent with the results obtained for the standard $M/G/1$ queue (see, for instance, Chapters 1 and 5 of Takács [20] and Kleinrock [14], respectively).

ACKNOWLEDGMENTS

The author is grateful to the Associate Editor and referees for their constructive comments on an earlier version of this paper. The preparation of this work has been done while visiting the Instituto Superior Técnico of the Technical University of Lisbon, from January to September 1997, under the Postdoctoral Complutense Program Abroad.

REFERENCES

[1] J. Abate and W. Whitt, The Fourier-series method for inverting transforms of probability distributions, Queueing Syst 10 (1992), 5–88.
[2] J.R. Artalejo and G.I. Falin, On the orbit characteristics of the $M/G/1$ retrial queue, Nav Res Logistics 43 (1996), 1147–1161.
[3] J.R. Artalejo and A. Gómez-Corral, Steady state solution of a single-server queue with linear request repeated, J Appl Probab 34 (1997), 223–233.
[4] 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 Syst 14 (1993), 275–292.
[5] G.L. Choudhury, D.M. Lucantoni, and W. Whitt, Multidimensional transform inversion with applications to the transient $M/G/1$ queue, Ann Appl Probab 4 (1994), 719–740.
[6] R.B. Cooper, Introduction to queueing theory, Edward Arnold, 1981.
[7] G.I. Falin, Single-line repeated orders queueing systems, Optimization 17 (1986), 649–677.
[8] G.I. Falin, A survey of retrial queues, Queueing Syst 7 (1990), 127–168.
[9] G.I. Falin and J.G.C. Templeton, Retrial queues, Chapman and Hall, New York, 1997.
[10] G. Fayolle, "A simple telephone exchange with delayed feedbacks," Teletraffic analysis and computer performance evaluation, O.J. Boxma, J.W. Cohen, and H.C. Tijms (Editors), Elsevier, Amsterdam, 1986, pp. 245–253.
[11] S.W. Fuhrmann and R.B. Cooper, Stochastic decomposition in the $M/G/1$ queue with generalized vacations, Oper Res 33 (1985), 1117–1129.
[12] V.A. Kapyrin, A study of the stationary characteristics of a queueing system with recurring demands, Cybernetics 13 (1977), 584–590.
[13] J. Keilson, J. Cozzolino, and H. Young, A service system with unfilled requests repeated, Oper Res 16 (1986), 1126–1137.
[14] L. Kleinrock, Queueing systems, vol. I: Theory, Wiley-Interscience, New York, 1975.

[15] V.G. Kulkarni and H.M. Liang, “Retrial queues revisited,” Frontiers in queueing. Models and applications in science and engineering, J.H. Dshalalow (Editor), CRC Press, Boca Raton, FL, 1997, pp. 19–34.

[16] M. Martin and J.R. Artalejo, Analysis of an $M/G/1$ queue with two types of impatient units, Adv Appl Probab 27 (1995), 840–861.

[17] M.F. Neuts and M.F. Ramalhoto, A service model in which the server is required to search for customers, J Appl Probab 21 (1984), 157–166.

[18] B. Pourbabai, A note on a $D/G/K$ loss system with retrials, J Appl Probab 27 (1990), 385–392.

[19] L.I. Sennot, P.A. Humblet, and R.L. Tweedie, Mean drifts and the non-ergodicity of Markov chains, Oper Res 31 (1983), 783–789.

[20] L. Takács, Introduction to the theory of queues, Oxford University Press, Oxford, 1962.

[21] T. Yang and J.G.C. Templeton, A survey on retrial queues, Queueing Syst 2 (1987), 201–233.

[22] T. Yang, M.J.M. Posner, J.G.C. Templeton, and H. Li, An approximation method for the $M/G/1$ retrial queue with general retrial times, Eur J Oper Res 76 (1994), 552–562.