An $M/M/1$ retrial queue with control policy and general retrial times

Bong Dae Choi *

Department of Mathematics, Korea Advanced Institute of Science and Technology, 373-1 Kusong-dong, Yusong-gu, Taejon, Korea

Kwang Kyu Park

Electronics and Telecommunications Research Institute, P. O. Box 8, Daejon, Korea

C.E.M. Pearce

Applied Mathematics Department, Adelaide University, G.P.O. Box 498, Adelaide S.A. 5001, Australia

Received 12 June 1992; revised 1 December 1992

We consider an $M/M/1$ retrial queueing system in which the retrial time has a general distribution and only the customer at the head of the queue is allowed to retry for service. We find a necessary and sufficient condition for ergodicity and, when this is satisfied, the generating function of the distribution of the number of customers in the queue and the Laplace transform of the waiting time distribution under steady-state conditions. The results agree with known results for special cases.

Keywords: Retrial queue; control policy; general retrial times; supplementary variable method.

1. Introduction

We consider an $M/M/1$ retrial queueing system with exogenous Poisson arrivals occurring at rate λ and customer service times which are independent and exponentially distributed with mean $1/\mu$. An arrival obtains service immediately if the server is idle and joins the queue (in accordance with a first- come, first-served discipline) if the server is busy. The control policy for access to the server from the retrial group is that only the customer at the head of the queue can retry for service. This is tried for after a random time. If at the instant of retrial the customer concerned finds the server busy, he returns to the head of the queue and repeats this procedure until he succeeds. We assume that the retrial times (the time intervals

* Supported by KOSEF 90-08-00-02.

between two consecutive retrials made by a customer) are independent and identically distributed with cumulative distribution function $S(\cdot)$, density function $R(\cdot)$ and mean r.

Fayolle [7] has investigated, as a telephone exchange model, an $M/M/1$ retrial queue with exponential retrial times and the same control policy except that retrial customers who find the server busy join the tail of the queue instead of the head. He found the queue size and sojourn time distributions. We note that as the queue size distribution is independent of queue discipline, his expression for the queue size distribution should agree with ours for the case of exponential retrial times. Farahmand [6] has investigated and $M/G/1$ retrial queue with exponential retrial times and the same control policy and obtained the queue length distribution. See also Choi and Park [1] who use a supplementary variable method. Falin [5] and Yang and Templeton [12] give comprehensive reviews of the retrial literature.

An examination of the literature reveals the remarkable fact that even retrial systems with all exponential components can be very complicated. Many have not been solved at all and those that have been do not always possess generating functions for the equilibrium distribution of state which are such that the component probabilities can be easily extracted. For example, a direct treatment of the classical retrial problem considered by Cohen [3] seems to necessitate the use of extended continued fractions (see Hanschke [8] and Pearce [10]). The literature on non-exponential retrial times is very sparse. Approximate methods have been used by Pourbabai [11] on the $G/M/S/0$ retrial queue and by Yang [13] on the $M/G/1$ retrial queue. To the best of the authors' knowledge the present work is the first giving exact analytic results for a retrial problem with generally distributed retrial times.

In section 2 we employ an imbedded Markov chain to derive a necessary and sufficient condition for ergodicity of the system. In section 3 we use a supplementary variable approach to find the generating function of the number of customers in the queue. This result is shown to coincide with a known one [7] in the case when retrial times are exponentially distributed. Section 4 contains a treatment of the queue length problem useful for algorithmic calculation. In section 5, we find the Laplace transform of the waiting time distribution. Finally, in section 6, we outline a generating function-free treatment of the waiting time.

2. Ergodicity

Let $N(t)$ represent the number of customers in the queue at time t. We write $\xi(t) = 1$ or 0 according as the server is busy or free at time t. For $N(t) > 0$ we denote by $\tilde{R}(t)$ the residual amount of the present retrial lifetime of the customer at the head of the queue.

The Markov process $Y(t) \equiv (N(t), \tilde{R}(t), \xi(t))$ is clearly regular, having bounded transition rates, and so almost surely only finitely many transitions of state occur in any finite time. We utilize an associated discrete-time Markov chain (X_n) in which the continuous-time process is observed just after the completion of each retrial. The state of (X_n) at each such time is the corresponding retrial queue size. As the underlying process is regular, it will be ergodic if and only if the Markov chain is ergodic. We shall derive the following result.

THEOREM 1

A necessary and sufficient condition for the ergodicity of the retrial system is that

$$\left[1 - \int_0^\infty e^{-(\lambda+\mu)x}\, dR(x)\right] \frac{\mu^2}{\lambda+\mu} > \lambda^2 r .$$

Proof

First we consider the one-step transition probabilities $(P_{i,j})$ for (X_n). For $j \geqslant i > 0$ and $k \in \{0,1\}$, let $\beta_{k,i,j}(t)$ represent the probability that $\xi(t) = k$ and $N(t) = j$, conditional on $\xi(0) = 1$, $N(0) = i$ and a retrial lifetime having begun at time 0 but not having finished by time t. Then

$$P_{i,j} = \int_0^\infty [\beta_{1,i,j}(x) + \beta_{0,i,j+1}(x)]\, dR(x) \quad (j \geqslant i > 0),$$

$$P_{i,i-1} = \int_0^\infty \beta_{0,i,i}(x)\, dR(x) \quad (i > 0),$$

$$P_{0,j} = \int_0^\infty [\beta_{1,1,j}(x) + \beta_{0,1,j+1}(x)]\, dR(x) \quad (j > 0),$$

$$P_{0,0} = \int_0^\infty \beta_{0,1,1}(x)\, dR(x).$$

For each $m \geqslant 0$,

$$\beta_{k,1+m,n+m}(x) = \beta_{k,1,n}(x) = \alpha_{k,n}(x), \quad \text{say},$$

so the transition probabilities thus have the form

$$P_{0,j} = P_{1,j} \quad (j \geqslant 0),$$

$$P_{i,j} = \begin{cases} f(j-i+1) & (j \geqslant i-1 \geqslant 0), \\ 0 & (j < i-1), \end{cases} \tag{2.1}$$

where

$$f(0) = \int_0^\infty \alpha_{0,1}(x)\, dR(x),$$

$$f(j) = \int_0^\infty [\alpha_{1,j}(x) + \alpha_{0,j+1}(x)]\, dR(x) \quad (j > 0).$$

The structure exhibited by (2.1) is just that of the transition probabilities in the standard $M/G/1$ system. A necessary and sufficient condition for ergodicity may be derived exactly as for that system to be

$$F'(1) < 1\,, \tag{2.2}$$

where

$$F(z) = \sum_{j=0}^{\infty} f(j) z^j = \int_0^{\infty} [A_1(z,x) + A_0(z,x)]\, \mathrm{d}S\,(x) \quad (|z| \leqslant 1)\,, \tag{2.3}$$

the functions A_k being defined by

$$A_0(z,x) := \sum_{j=0}^{\infty} \alpha_{0,j+1}(x) z^j \quad (|z| \leqslant 1)\,,$$
$$A_1(z,x) := \sum_{j=1}^{\infty} \alpha_{1,j}(x) z^j \quad (|z| \leqslant 1)\,.$$

The quantities $\alpha_{k,n}(t)$ are given by the forward differential equations

$$\alpha'_{0,n}(t) = -\lambda \alpha_{0,n}(t) + \mu \alpha_{1,n}(t) \quad (n > 0)\,,$$
$$\alpha'_{1,n}(t) = -(\lambda + \mu)\alpha_{1,n}(t) + \lambda \alpha_{0,n}(t) + \lambda(1 - \delta_{n,1})\alpha_{1,n-1}(t) \quad (n > 0)\,,$$

subject to the initial conditions $\alpha_{0,n}(0) = 0$ and $\alpha_{1,n}(0) = \delta_{n,1}$. These equations may be cast in terms of the Laplace transforms

$$\alpha^*_{k,n}(s) := \int_0^{\infty} \mathrm{e}^{-st} \alpha_{k,n}(t)\, \mathrm{d}t \quad (\Re s \geqslant 0)$$

as

$$(s + \lambda)\alpha^*_{0,n}(s) = \mu \alpha^*_{1,n}(s) \quad (n > 0)\,,$$
$$(s + \lambda + \mu)\alpha^*_{1,n}(s) - \delta_{n,1} = \lambda \alpha^*_{0,n}(s) + \lambda(1 - \delta_{n,1})\alpha^*_{1,n-1}(s) \quad (n > 0)\,.$$

Finally, if $A^*_k(z,s)$ represents the Laplace transform of $A_k(z,t)$, we have

$$z(s + \lambda)A^*_0(z,s) = \mu A^*_1(z,s)\,,$$
$$[s + \lambda(1 - z) + \mu]A^*_1(z,s) = z + \lambda z A^*_0(z,s)\,. \tag{2.4}$$

On substituting $z = 1$ in these equations, we derive

$$A^*_0(1,s) = \frac{\mu}{s(s + \lambda + \mu)}\,,$$
$$A^*_1(1,s) = \frac{s + \lambda}{s(s + \lambda + \mu)}\,.$$

Also, if $A^{*\prime}_k(1,s)$ denotes $\partial/\partial z A^*_k(z,s)|_{z=1}$, we have on differentiation of (2.4) that

$$(s+\lambda)[A_0^*(1,s)+A_0^{*\prime}(1,s)] = \mu A_1^{*\prime}(1,s),$$
$$(s+\mu)A_1^{*\prime}(1,s) - \lambda A_1^*(1,s) = 1 + \lambda[A_0^*(1,s)+A_0^{*\prime}(1,s)],$$

whence we find

$$A_1^{*\prime}(1,s) = \frac{(s+\lambda)[\lambda A_1^*(1,s)+1]}{s(s+\lambda+\mu)}$$

and thus

$$(s+\lambda)\sum_{k=0,1} A_k^{*\prime}(1,s) = (s+\lambda+\mu)A_1^{*\prime}(1,s) - (s+\lambda)A_0^*(1,s) = \frac{(s+\lambda)^3}{s^2(s+\lambda+\mu)},$$

so that on inverting the transform we obtain

$$\partial/\partial z \sum_{k=0,1} A_k(z,t)|_{z=1} = \frac{\lambda}{\lambda+\mu} + \frac{\lambda\mu}{(\lambda+\mu)^2} + \frac{\lambda^2 t}{\lambda+\mu} + \frac{\mu^2}{(\lambda+\mu)^2}\mathrm{e}^{-(\lambda+\mu)t}.$$

From (2.3), condition (2.2) thus reads

$$\frac{\lambda}{\lambda+\mu} + \frac{\lambda\mu}{(\lambda+\mu)^2} + \frac{\lambda^2}{\lambda+\mu}r + \frac{\mu^2}{(\lambda+\mu)^2}\int_0^\infty \mathrm{e}^{-(\lambda+\mu)x}\,\mathrm{d}S(x) < 1,$$

which on simplifications gives theorem 1. □

3. Queue size distribution

Let the supplementary variable $\tilde{R}(t)$ denote the remaining retrial time of the customer at the head of the queue when the queue is not empty at time t. We define

$$p_0(t) = P(N(t)=0, \xi(t)=1),$$
$$q_0(t) = P(N(t)=0, \xi(t)=0),$$
$$p_i(t,x)dx = P(N(t)=i, \tilde{R}(t)\in[x,x+dx), \xi(t)=1), \quad i\geqslant 1,$$
$$q_i(t,x)dx = P(N(t)=i, \tilde{R}(t)\in[x,x+dx), \xi(t)=0), \quad i\geqslant 1.$$

By considering transitions of the process between times t and $t+\Delta t$ and letting $\Delta t\to 0$, we derive (as in [1,9]) the system of forward equations

$$\frac{\mathrm{d}}{\mathrm{d}t}p_0(t) = -(\lambda+\mu)p_0(t) + \lambda q_0(t) + q_1(t,0), \tag{3.1a}$$

$$\left(\frac{\partial}{\partial t}-\frac{\partial}{\partial x}\right)p_1(t,x) = -(\lambda+\mu)p_1(t,x) + \lambda r(x)p_0(t) + r(x)p_1(t,0) + \lambda q_1(t,x) + r(x)q_2(t,0)\,, \tag{3.1b}$$

$$\left(\frac{\partial}{\partial t}-\frac{\partial}{\partial x}\right)p_i(t,x) = -(\lambda+\mu)p_i(t,x) + \lambda p_{i-1}(t,x) + r(x)p_i(t,0) + \lambda q_i(t,x) + r(x)q_{i+1}(t,0)\,, \quad i\geqslant 2\,, \tag{3.1c}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}q_0(t) = -\lambda q_0(t) + \mu p_0(t)\,, \tag{3.1d}$$

$$\left(\frac{\partial}{\partial t}-\frac{\partial}{\partial x}\right)q_i(t,x) = -\lambda q_i(t,x) + \mu p_i(t,x)\,, \quad i\geqslant 1\,, \tag{3.1e}$$

together with the normalizing equation

$$p_0(t) + q_0(t) + \sum_{i=1}^{\infty}\int_0^{\infty}[p_i(t,x) + q_i(t,x)]\,\mathrm{d}x = 1\,.$$

Since we are interested in the steady-state behaviour of the system, let us assume that the condition for the system to be stable as $t\to\infty$ is satisfied. We write $p_0 = \lim_{t\to\infty}p_0(t), p_i(x) = \lim_{t\to\infty}p_i(t,x)$ and similarly for the q's and derive from the forward equations above that

$$(\lambda+\mu)p_0 = \lambda q_0 + q_1(0)\,, \tag{3.2a}$$

$$-\frac{\mathrm{d}}{\mathrm{d}x}p_1(x) = -(\lambda+\mu)p_1(x) + \lambda r(x)p_0 + r(x)p_1(0) + \lambda q_1(x) + r(x)q_2(0)\,, \tag{3.2b}$$

$$-\frac{\mathrm{d}}{\mathrm{d}x}p_i(x) = -(\lambda+\mu)p_i(x) + \lambda p_{i-1}(x) + r(x)p_i(0) + \lambda q_i(x) + r(x)q_{i+1}(0)\,, \quad i\geqslant 2\,, \tag{3.2c}$$

$$\lambda q_0 = \mu p_0\,, \tag{3.2d}$$

$$-\frac{\mathrm{d}}{\mathrm{d}x}q_i(x) = -\lambda q_i(x) + \mu p_i(x)\,, \quad i\geqslant 1\,. \tag{3.2e}$$

The normalizing equation becomes

$$p_0 + q_0 + \sum_{i=1}^{\infty}\int_0^{\infty}[p_i(x) + q_i(x)]\,\mathrm{d}x = 1\,.$$

From (3.2a) and (3.2d) we obtain that

$$\lambda p_0 = q_1(0) .$$

With notation along the lines of the previous section, we employ $p_i^*(\theta)$ for $i \geqslant 1$ to denote the Laplace transform of p_i (with $\theta \geqslant 0$) and similarly for $q_i^*(\theta)$. The quantity

$$p_i^*(0) = P(N = i, \xi = 1)$$

is then the steady-state probability that there are i customers in the queue and the server is busy.

For z satisfying $-1 \leqslant z \leqslant 1$, define the generating functions

$$P^*(\theta, z) = \sum_{i=1}^{\infty} p_i^*(\theta) z^i , \quad P(0, z) = \sum_{i=1}^{\infty} p_i(0) z^i ,$$

and corresponding quantities $Q^*(\theta, z)$, $Q(0, z)$. We observe that $P^*(0, z) + Q^*(0, z) + p_0 + q_0$ is the generating function for the distribution of the number of customers in the retrial group at an arbitrary time point in the steady state.

On taking Laplace transforms and forming generating functions from (3.1b, c, e) we derive the basic equations

$$(\theta - \lambda) Q^*(\theta, z) + \mu P^*(\theta, z) = Q(0, z) , \tag{3.3}$$

$$\lambda Q^*(\theta, z) + (\theta - \lambda - \mu + \lambda z) P^*(\theta, z) = f(\theta, z) , \tag{3.4}$$

where

$$f(\theta, z) = P(0, z) + r^*(\theta) \left\{ \lambda (1 - z) p_0 - P(0, z) - \frac{Q(0, z)}{z} \right\} .$$

We wish to solve these equations for P^* and Q^*. Let $\theta_1(z), \theta_2(z)$ be, respectively, the greater and lesser zeros of the quadratic polynomial in θ

$$D(\theta, z) = (\theta - \lambda)(\theta - \lambda - \mu + \lambda z) - \lambda \mu .$$

We can easily check that both zeros are nonnegative for $-1 \leqslant z \leqslant 1$ and that $\theta_1(1) = \lambda + \mu$ and $\theta_2(1) = 0$. For convenience we henceforth omit the argument z in $\theta_i(z)$ $(i = 1, 2)$ except where it is required to avoid ambiguity.

By applying Cramér's rule to (3.3) and (3.4) we have

$$D(\theta, z) Q^*(\theta, z) = Q(0, z)(\theta - \lambda - \mu + \lambda z) - \mu f(\theta, z) , \tag{3.5}$$

$$D(\theta, z) P^*(\theta, z) = (\theta - \lambda) f(\theta, z) - \lambda Q(0, z) . \tag{3.6}$$

Letting $\theta = 0$ in (3.5) and (3.6) yields

$$Q^*(0, z) + q_0 = \lambda^{-2}(\mu / z - \lambda) Q(0, z) , \tag{3.7}$$

$$P^*(0, z) + p_0 = (\lambda z)^{-1} Q(0, z) . \tag{3.8}$$

Since $P^*(\theta, z)$ has a finite value for any $\theta \geqslant 0$ and $-1 \leqslant z \leqslant 1$ and $D(\theta, z)$ vanishes at

θ_i $(i = 1, 2)$, the right-hand sides of (3.5) and (3.6) must vanish at $\theta = \theta_1, \theta_2$. Thus we have

$$(\theta_i - \lambda)[1 - r^*(\theta_i)]P(0,z) - \left[\lambda + (\theta_i - \lambda)\frac{r^*(\theta_i)}{z}\right]Q(0,z) \\ = \lambda(\theta_i - \lambda)r^*(\theta_i)(z-1)p_0$$

for $i = 1, 2$. On applying Cramér's rule to these two equations we have

$$P(0,z) = \lambda p_0(z-1)K(z)^{-1}[\lambda(\theta_2 - \lambda)r^*(\theta_2) - \lambda(\theta_1 - \lambda)r^*(\theta_1)], \tag{3.9}$$

$$Q(0,z) = \lambda p_0(z-1)K(z)^{-1}(\theta_1 - \lambda)(\theta_2 - \lambda)[r^*(\theta_2) - r^*(\theta_1)], \tag{3.10}$$

where

$$K(z) = \begin{vmatrix} (\theta_1 - \lambda)(1 - r^*(\theta_1)) & -[\lambda + (\theta_1 - \lambda)r^*(\theta_1)/z] \\ (\theta_2 - \lambda)(1 - r^*(\theta_2)) & -[\lambda + (\theta_2 - \lambda)r^*(\theta_2)/z] \end{vmatrix}.$$

Letting $z \to 1$ in (3.10) yields

$$Q(0,1) = \frac{\lambda\mu(1 - r^*(\lambda + \mu))p_0}{(1 - r^*(\lambda + \mu))\dfrac{\mu^2}{\lambda + \mu} - \lambda^2 r},$$

and letting $z \to 1$ in (3.7) and (3.8) gives

$$Q^*(0,1) + q_0 = \lambda^{-2}(\mu - \lambda)Q(0,1) \quad \text{and} \quad P^*(0,1) + p_0 = \lambda^{-1}Q(0,1).$$

Since $P^*(0,1) + Q^*(0,1) + p_0 + q_0 = 1$ and $p_0 = \lambda q_0/\mu$, we thus have

$$p_0 = \frac{(1 - r^*(\lambda + \mu))\dfrac{\lambda\mu^2}{\lambda + \mu} - \lambda^3 r}{\mu^2(1 - r^*(\lambda + \mu))},$$

$$q_0 = \frac{(1 - r^*(\lambda + \mu))\dfrac{\mu^2}{\lambda + \mu} - \lambda^2 r}{\mu(1 - r^*(\lambda + \mu))}. \tag{3.11}$$

As we saw in section 2, the positivity of the numerators is a necessary and sufficient condition for system stability.

From (3.10) we derive

$$Q(0,z) = \frac{\lambda\mu(z-1)[r^*(\theta_1) - r^*(\theta_2)]p_0 z}{\sum_{i=1,2}(-1)^i[z\theta_i(1 - r^*(\theta_i)) + (\lambda z + \mu)r^*(\theta_i)]}, \tag{3.12}$$

where we have used $(\theta_1 - \lambda)(\theta_2 - \lambda) = -\lambda\mu$. From (3.9) and (3.10), $P(0,z)$ is now known from

$$P(0,z) = Q(0,z)\left[\frac{\lambda}{\mu} - \frac{\theta_2 r^*(\theta_2) - \theta_1 r^*(\theta_1)}{\mu[r^*(\theta_2) - r^*(\theta_1)]}\right],$$

so we have $f(\theta, z)$ from its definition in terms of $P(0,z)$, $Q(0,z)$ and p_0. Addition of (3.5) and (3.6) gives

$$P^*(\theta,z)+Q^*(\theta,z)=\frac{(\theta-\lambda-\mu)[Q(0,z)+f(\theta,z)]-\lambda(1-z)Q(0,z)}{(\theta-\lambda)(\theta-\lambda-\mu+\lambda z)-\lambda\mu}. \quad (3.13)$$

We thus have obtained the following result.

THEOREM 2

(a) The generating functions for the distributions of the numbers of customers in the retrial queue when the server is idle or busy are given, respectively, by the right-hand sides of (3.7) and (3.8), where $Q(0,z)$ is given by (3.12);

(b) the joint distribution of the number of customers in the queue and the remaining retrial time of the customer at the head of the queue is given through its transform $P^*(\theta,z)+Q^*(\theta,z)$ by (3.13).

Exponential retrial times

When retrial times have an exponential distribution with mean r, our model becomes that of Fayolle [7] with a different service discipline but the same queue size distribution. We verify that our result reduces to the known one.

Substitution of $r^*(\theta)=(1/r)/(\theta+1/r)$ into (3.11) yields

$$q_0=1-(1+\lambda r)\rho,$$
$$p_0=\rho[1-(1+\lambda r)\rho],$$

where $\rho=\lambda/\mu$. Also $\theta_1\theta_2=\lambda^2(1-z)$ and $\theta_1(\theta_1-\lambda)-\theta_2(\theta_2-\lambda)=(\lambda+\mu-\lambda z)\times(\theta_1-\theta_2)$, so that

$$K(z)=\frac{(z-1)(\theta_2-\theta_1)}{z(\theta_1+1/r)(\theta_2+1/r)}\left[\frac{\lambda}{r}(\lambda+\mu-\lambda z)-\lambda^2(\lambda z+\mu)\right]$$

and the term following $K(z)^{-1}$ in (3.10) is equal to

$$\frac{\lambda\mu}{r}\frac{\theta_2-\theta_1}{(\theta_1+1/r)(\theta_2+1/r)}.$$

Substitution of the values into (3.10) gives

$$Q(0,z)=\frac{\rho}{r}\frac{[1-(1+\lambda r)\rho]z}{1-(1+\lambda r)\rho z}.$$

Thus from theorem 2 we have

$$E(z^N;\xi=0)=Q^*(0,z)+q_0=\frac{(1-\rho z)[1-(1+\lambda r)\rho]}{1-(1+\lambda r)\rho z},$$
$$E(z^N;\xi=1)=P^*(0,z)+p_0=\rho\frac{1-(1+\lambda r)\rho}{1-(1+\lambda r)\rho z},$$

which is proposition 1 of Fayolle [7].

4. Determination of the queue size probabilities

In this section we derive, under the stability assumption, formulae suitable for algorithmic calculation of the quantities $p_i(x), q_i(x)$ $(i \geqslant 1)$. Our derivation reveals some symmetries between the p- and q-quantities less apparent in the formulation of the previous section, and the equations are rendered more compact if we vary the notation of section 3 and for $i \geqslant 1$ set

$$\pi_{k,j}(x) = \begin{cases} q_j(x) & \text{for } k = 0, \\ p_j(x) & \text{for } k = 1, \end{cases}$$

and define

$$\pi_{k,0} = \begin{cases} q_0 & \text{for } k = 0, \\ p_0 & \text{for } k = 1. \end{cases}$$

The analysis is based on the imbedded Markov chain utilized in section 2. If we denote the equilibrium distribution on the chain by $(\phi_j)_0^\infty$, then following the standard theory of the $M/G/1$ queue verbatim we derive the generating function

$$\sum_{j=0}^{\infty} \phi_j z^j = [1 - F'(1)] \frac{(z-1)F(z)}{z - F(z)} .$$

An algorithmic procedure for the calculation of the probabilities ϕ_j is given by Çinlar as theorem 5.20 in [2]. Let $y_j = 1 - \sum_{i=0}^{j} f_i$. Then

$$\phi_0 = 1 - F'(1) ,$$
$$\phi_1 / \phi_0 = y_0 / f_0$$

and for any $j > 1$

$$\phi_j / \phi_0 = \sum_{i=1}^{j-1} f_0^{-i-1} \sum y_{\epsilon_1} y_{\epsilon_2} \dots y_{\epsilon_i} ,$$

where the inner summation is over all i-tuples of positive integers with

$$\epsilon_1 + \epsilon_2 + \dots \epsilon_i = j - 1 .$$

For $k = 1, 2$, let $\gamma_k(t)$ be the probability that $\xi(t) = k$ and no retrial lifetime has begun since time 0, given a retrial lifetime ended at time 0 and emptied the retrial queue. Then

$$\gamma_0'(t) = \lambda \gamma_0(t) + \mu \gamma_1(t) ,$$
$$\gamma_1'(t) = -(\lambda + \mu) \gamma_1(t) + \lambda \gamma_0(t)$$

with initial conditions $\gamma_k(0) = \delta_{k,1}$.

In terms of Laplace transforms, these forward equations may be written as

$$(s+\lambda)\gamma_0^*(s) = \mu\gamma_1^*(s)\,,$$
$$(s+\lambda+\mu)\gamma_1^*(s) = 1 + \lambda\gamma_0^*(s)\,,$$

so that

$$\gamma_0^*(s) = \frac{\mu}{s(s+2\lambda+\mu)+\lambda^2}\,,$$
$$\gamma_1^*(s) = \frac{s+\lambda}{s(s+2\lambda+\mu)+\lambda^2}\,.$$

The probability density for an empty period of the retrial queue having length t is $\lambda\gamma_1(t)$, so that the mean length of empty periods is

$$\lambda\int_0^\infty t\gamma_1(t)\,\mathrm{d}t = -\lambda\gamma_1^{*\prime}(0) = \mu/\lambda^2\,.$$

Hence the mean time between the ends of successive busy periods for the retrial queue is $(p_0+q_0)^{-1}\mu/\lambda^2$. Such time points occur in a renewal stream at rate $(p_0+q_0)\lambda^2/\mu$. Chain points at which the state is j will occur in a renewal stream. The rate for these will be

$$\psi_j = (p_0+q_0)(\lambda^2/\mu)\phi_j/\phi_0 \quad (j\geqslant 0)\,,$$

which is an algorithmically calculable quantity since p_0+q_0 is given from section 3 by

$$p_0+q_0 = 1 - \frac{(\lambda+\mu)\lambda^2 r}{\mu^2[1-r^*(\lambda+\mu)]}\,.$$

We determine $\pi_{k,j}(x)$ by use of the theorem of total probability, conditioning on the elapsed time since the last event epoch of the imbedded chain and the state of the chain at that epoch. By renewal theory we have

$$\pi_{k,j}(x) = \sum_{l=1}^{j}\psi_l\int_0^\infty \beta_{k,l,j}(t)r(t+x)\,\mathrm{d}t$$
$$+\psi_0\int_0^\infty \gamma_1(\tau)\lambda\,\mathrm{d}\tau\int_0^\infty \beta_{k,1,j}(t)r(t+x)\,\mathrm{d}t \quad (j>0)\,.$$

Since

$$\int_0^\infty \gamma_1(\tau)\lambda\,\mathrm{d}\tau = \lambda\gamma_1^*(0) = 1\,,$$

this equation simplifies to

$$\pi_{k,j}(x) = \sum_{l=1}^{j}\psi_l\int_0^\infty \alpha_{k,j-l+1}(t)r(t+x)\,\mathrm{d}t$$
$$+\psi_0\int_0^\infty \alpha_{k,j}(t)r(t+x)\,\mathrm{d}t \quad (j>0)\,.$$

It remains to determine the quantities $\alpha_{k,j}(t)$. From (2.4) we have

$$A_0^*(z,s) = \frac{\mu}{(s+\lambda)[s+\lambda(1-z)+\mu]-\lambda\mu},$$
$$A_1^*(z,s) = \frac{z(s+\lambda)}{(s+\lambda)[s+\lambda(1-z)+\mu]-\lambda\mu},$$

whence we derive

$$\begin{aligned}\alpha_{0,j}^*(s) &= \frac{\mu[\lambda(s+\lambda)]^{j-1}}{[s(s+2\lambda+\mu)+\lambda^2]^j} \quad (j>0),\\ \alpha_{1,j}^*(s) &= \frac{\lambda^{j-1}(s+\lambda)^j}{[s(s+2\lambda+\mu)+\lambda^2]^j} \quad (j>0).\end{aligned} \tag{4.1}$$

These transforms can be inverted by use of formula (5.2.17) of Erdélyi et al., which gives $\sum_{m=1}^{j}\lambda_m s^{j-m}(s+a)^{-j}$ as the Laplace transform of

$$e^{-at}\sum_{m=0}^{j-1}\frac{t^m}{m!}\sum_{n=0}^{m}\binom{j-m-1+n}{n}(-1)^n\lambda_{m+1-n}a^n, \tag{4.2}$$

where we interpret $\binom{0}{0}=1$. If we make the factorization

$$s(s+2\lambda+\mu)+\lambda^2 = (s+v_1)(s+v_2),$$

then v_1, v_2 are clearly real and nonnegative. From (4.2), $(s+\lambda)^{j-1}/(s+v_1)^j$ is the Laplace transform of

$$e^{-v_1 t}\sum_{m=0}^{j-1}\frac{[t(\lambda-v_1)]^m}{m!}\binom{j-1}{m}.$$

Since $(s+v_2)^{-j}$ is the Laplace transform of

$$e^{-v_2 t}t^{j-1}/(j-1)!,$$

we may invert the first relation of (4.1) by the convolution theorem as

$$\alpha_{0,j}(t) = \mu\int_0^t e^{-v_2(t-u)}\frac{(\lambda(t-u))^{j-1}}{(j-1)!}e^{-v_1 u}\sum_{m=0}^{j-1}\frac{((\lambda-v_1)u)^m}{m!}\binom{j-1}{m}\,du \quad (j>0).$$

Similarly

$$\begin{aligned}\alpha_{1,j}(t) = \int_0^t e^{-v_2(t-u)}&\left[\lambda\frac{(\lambda(t-u))^{j-2}}{(j-2)!}+(\lambda-v_2)\frac{(\lambda(t-u))^{j-1}}{(j-1)!}\right]\\ &\times e^{-v_1 u}\sum_{m=0}^{j-1}\frac{((\lambda-v_1)u)^m}{m!}\binom{j-1}{m}\,du \quad (j>0)\end{aligned}$$

and we are finished.

5. Waiting time distribution

Let T be the random variable representing the waiting time of a tagged customer in the queue. By conditioning the state of the system seen by such a customer at his external arrival time, we have

$$E(e^{-sT}) = P(\xi = 0) + E(e^{-sT}|N = 0, \xi = 1)p_0 + \sum_{n=1}^{\infty} \int_0^{\infty} E(e^{-sT}|N = n, \tilde{R} = x, \xi = 1)p_n(x)\, dx. \quad (5.1)$$

We observe that $E(e^{-sT}|N = n, \tilde{R} = x, \xi = 1)$ (for $n \geqslant 1$) is the Laplace transform of the waiting time in the queue of the arriving tagged customer when there are n customers in the retrial group, the remaining retrial time of the customer at the head of the queue is x and the server is busy. This transform is determined by modifying the original time-dependent $M/M/1$ retrial queue differential-difference equations (3.1) with initial conditions $N(0) = n+1, \tilde{R}(0) = x, \xi(0) = 1$ and the input policy modified so that external arrivals who find the server busy are lost. With these changes $q_1(t, 0)$ will be the required probability density function of the conditional waiting time and $\int_0^{\infty} e^{-st} q_1(t, 0)\, dt$ (later denoted by $w^*_{n+1}(s, x)$) will be $E(e^{-sT}|N = n, \tilde{R} = x, \xi = 1)$. The main object of the following argument is to find $\int e^{-st} q_1(t, 0)\, dt$.

With the new input policy, eqs. (3.1b, c, e) become

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial y}\right)p_{n+1}(t, y) = -\mu p_{n+1}(t, y) + r(y)p_{n+1}(t, 0) + \lambda q_{n+1}(t, y), \quad (5.2a)$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial y}\right)p_i(t, y) = -\mu p_i(t, y) + r(y)p_i(t, 0) + \lambda q_i(t, y) + r(y)q_{i+1}(t, 0), \quad 1 \leqslant i \leqslant n, \quad (5.2b)$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial y}\right)q_i(t, y) = -\lambda q_i(t, y) + \mu p_i(t, y), \quad 1 \leqslant i \leqslant n+1, \quad (5.2c)$$

with the initial conditions

$$p_i(0, y) = \delta_x(y)\delta_{i,n+1} \quad \text{and} \quad q_i(0, y) = 0. \quad (5.3)$$

We solve these differential-difference equations *via* the transforms

$$P^*(s, \theta, z) = \sum_{i=1}^{n+1} z^i \int_0^{\infty} \int_0^{\infty} \exp(-st - \theta y)p_i(t, y)\, dt\, dy,$$

$$Q^*(s, \theta, z) = \sum_{i=1}^{n+1} z^i \int_0^{\infty} \int_0^{\infty} \exp(-st - \theta y)q_i(t, y)\, dt\, dy,$$

$$U(s,z)=\sum_{i=1}^{n+1} z^i \int_0^\infty \exp(-st)p_i(t,0)\ \mathrm{d}t\,,$$

$$V(s,z)=\sum_{i=1}^{n+1} z^i \int_0^\infty \exp(-st)q_i(t,0)\ \mathrm{d}t\,.$$

First, forming transforms on (5.2) gives

$$\begin{aligned}&(\theta-s-\mu)P^*(s,\theta,z)+\lambda Q^*(s,\theta,z)=f(s,\theta,z)\,,\\ &\mu P^*(s,\theta,z)+(\theta-s-\lambda)Q^*(s,\theta,z)=V(s,z)\,,\end{aligned}\tag{5.4}$$

where

$$f(s,\theta,z)=U(s,z)+r^*(\theta)[w^*_{n+1}(s,x)-U(s,z)-V(s,z)/z]-\mathrm{e}^{-\theta x}z^{n+1}$$

and we set

$$w^*_{n+1}(s,x)=\int_0^\infty \mathrm{e}^{-st}q_1(t,0)\ \mathrm{d}t\,.$$

An application of Cramér's rule to (5.4) provides

$$\begin{aligned}&G(\theta,s)P^*(s,\theta,z)=f(s,\theta,z)(\theta-s-\lambda)-\lambda V(s,z)\,,\\ &G(\theta,s)Q^*(s,\theta,z)=(\theta-s-\mu)V(s,z)-\mu f(s,\theta,z)\,,\end{aligned}\tag{5.5}$$

where $G(\theta,s)=(\theta-s)(\theta-s-\lambda-\mu)$. Since $G(\theta,s)=0$ at $\theta=s$ and $\theta=s+\lambda+\mu$, the right-hand sides in (5.5) must also vanish there, whence we have

$$U(s,z)+r^*(s)[w^*_{n+1}(s,x)-U(s,z)-V(s,z)/z]-\mathrm{e}^{-sx}z^{n+1}+V(s,z)=0\,,$$

$$\begin{aligned}&\mu U(s,z)-\lambda V(s,z)+\mu r^*(s+\lambda+\mu)[w^*_{n+1}(s,x)-U(s,z)-V(s,z)/z]\\ &\quad-\mu\mathrm{e}^{-(s+\lambda+\mu)x}z^{n+1}=0\,.\end{aligned}$$

An application of Cramér's rule to these two relations gives

$$U(s,z)=L(s,z)/M(s,z)\,,$$

where

$$L(s,z)=\begin{vmatrix}\mathrm{e}^{-sx}z^{n+1}-r^*(s)w^*_{n+1}(s,x) & 1-r^*(s)/z\\ \mu\mathrm{e}^{-(s+\lambda+\mu)x}z^{n+1}-w^*_{n+1}(s,x)r^*(s+\lambda+\mu)\mu & -\lambda-\mu r^*(s+\lambda+\mu z^{-1})\end{vmatrix}\,,$$

$$M(s,z)=-\lambda-\mu+r^*(s)[\lambda+\mu/z]+\mu r^*(s+\lambda+\mu)[1-z^{-1}]\,.$$

We observe that $M(s,z)=0$ when

$$z=z(s)=\frac{\mu[r^*(s)-r^*(s+\lambda+\mu)]}{\lambda+\mu-\lambda r^*(s)-\mu r^*(s+\lambda+\mu)}$$

and $|z(s)|\leqslant 1$ for $s\geqslant 0$. Hence $L(s,z)=0$ at $z=z(s)$ too, that is,

$$w_{n+1}^*(s,x) = \frac{e^{-sx}[\lambda + \mu r^*(s+\lambda+\mu)/z(s)] + \mu e^{-(s+\lambda+\mu)x}[1 - r^*(s)/z(s)]}{\lambda r^*(s) + \mu r^*(s+\lambda+\mu)} z(s)^{n+1} .$$

Finally, since

$$E(e^{-sT}|N=0,\xi=1) = \int_0^\infty w_1^*(s,x)r(x)\,dx ,$$
$$E(e^{-sT}|N=n,\tilde{R}=x,\xi=1) = w_{n+1}^*(s,x) ,$$

we have from (5.1) that

$$\begin{aligned} E(e^{-sT} &= q_0 + Q^*(0,1) + p_0 z(s) \\ &+ [[\lambda + \mu r^*(s+\lambda+\mu)/z(s)]P^*(s,z(s)) + \mu[1 - r^*(s)/z(s)] \\ &\times P^*(s+\lambda+\mu, z(s))z(s)]/[\lambda r^*(s) + \mu r^*(s+\lambda+\mu)] \quad . \end{aligned} \tag{5.6}$$

Thus we have obtained the following.

THEOREM 3

The Laplace transform of the steady-state distribution of the waiting time in the queue is given by (5.6).

6. Algorithmic determination of waiting times

The determination of the distribution of the waiting time in the equilibrium regime may also be determined in the spirit of section 4 without the use of the generating function of the queue length distribution.

For $n>0$, let $V_{k,n}(x,t)$ be the probability density for the waiting time of an arriving customer being t, given that the customer finds $\xi = k, N = n$ and $\tilde{R} = x$. We define $V_{k,0}(x,t)$ to be the corresponding quantity for a customer at the head of the retrial queue when $\tilde{R} = x$. The unconditional waiting time of a customer arriving in the steady state then has the density function

$$w(t) = q_0\delta(t) + p_0 \int_0^t V_{1,0}(x,t)r(x)\,dx + \int_0^t \sum_{n=1}^\infty \pi_{1,n}(x)V_{1,n}(x,t)\,dx . \tag{6.1}$$

In this formula only the quantities $V_{1,n}$ remain to be determined. We outline briefly below how this can be done.

The backward Kolmogorov differential equations governing $V_{0,n}$, $V_{1,n}$ are

$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right) V_{0,n}(x,t) = -\lambda V_{0,n}(x,t) + \lambda V_{1,n}(x,t) , \tag{6.2}$$

$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right) V_{1,n}(x,t) = -\mu V_{1,n}(x,t) + \mu V_{0,n}(x,t) , \tag{6.3}$$

subject to boundary conditions

$$V_{0,n}(0,t) = \int_0^\infty V_{1,n-1}(x,t)r(x)\,\mathrm{d}x \quad (n>0)\,,$$

$$V_{0,0}(0,t) = \delta(t)\,,$$

$$V_{1,n}(0,t) = \int_0^\infty V_{1,n}(x,t)r(x)\,\mathrm{d}x \quad (n\geqslant 0)\,,$$

and initial conditions

$$V_{0,n}(x,0) = \delta(x)\delta_{n,0}\,,$$
$$V_{1,n}(x,0) = 0\,.$$

From (6.2) and (6.3)

$$\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial t}+\mu+\lambda\right)V_{1,n}(x,t) = 0\,,$$

so that

$$\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial t}+\mu+\lambda\right)V_{1,n}(x,t) = (\lambda+\mu)h_n(t-x)\,, \quad \text{say}\,, \tag{6.4}$$

for h_n some differentiable function. Also by (6.3)

$$\lambda V_{1,n}(x,t) + \mu V_{0,n}(x,t) = (\lambda+\mu)h_n(t-x)\,. \tag{6.5}$$

Put

$$V_{1,n}(x,t) = Z_n(y,x)\,, \tag{6.6}$$

where $y = t - x$. Then by (6.4)

$$\frac{\partial}{\partial x}Z_n(y,x) + (\lambda+\mu)Z_n(y,x) = (\lambda+\mu)h_n(y)\,,$$

whence

$$Z_n(y,x) = h_n(y) + \mu g_n(y)\mathrm{e}^{-(\lambda+\mu)x}\,,$$

where g_n is an arbitrary function of integration. From (6.5) and (6.6) we thus have the parametric solution

$$V_{1,n}(x,t) = h_n(t-x) + \mu g_n(t-x)\mathrm{e}^{-(\lambda+\mu)x}\,,$$

$$V_{0,n}(x,t) = h_n(t-x) - \lambda g_n(t-x)\mathrm{e}^{-(\lambda+\mu)x}\,, \tag{6.7}$$

From the initial conditions, we see that each g_n and h_n has nonnegative support.

Finally, the boundary conditions yield

$$h_n^*(s) - \lambda g_n^*(s) = h_{n-1}^*(s) r^*(s) + \mu g_{n-1}^*(s) r^*(s+\lambda+\mu) \quad (n>0)\,,$$

$$h_0^*(s) - \lambda g_0^*(s) = 1\,,$$

$$h_n^*(s) + \mu g_n^*(s) = h_{n-1}^*(s) r^*(s) + \mu g_n^*(s) r^*(s+\lambda+\mu) \quad (n\geqslant 0)\,,$$

whence

$$h_n^*(s) = \left[1 - \frac{1-r^*(s)}{1-r^*(s+\lambda+\mu)}\right]^n \left[1 - \frac{\lambda}{\mu}\frac{1-r^*(s)}{1-r^*(s+\lambda+\mu)}\right]^{-n-1} \quad (n\geqslant 0)\,,$$

$$g_n^*(s) = -\mu^{-1}\frac{1-r^*(s)}{1-r^*(s+\lambda+\mu)}\left[1 - \frac{1-r^*(s)}{1-r^*(s+\lambda+\mu)}\right]^n \times \left[1 - \frac{\lambda}{\mu}\frac{1-r^*(s)}{1-r^*(s+\lambda+\mu)}\right]^{-n-1} \quad (n\geqslant 0)\,.$$

An explicit expression for w may now be found by substitution from (6.7) into (6.1). If

$$\sigma_n(x) = \pi_{1,n}(x) e^{-(\lambda+\mu)x}\,,$$

then we derive

$$w(t) = q_0\delta(t) + p_0(h_0 * r)(t) + \sum_{n=1}^{\infty}(\pi_{1,n} * h_n)(t) + \mu\sum_{n=1}^{\infty}(\sigma_n * g_n)(t)\,.$$

References

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

[2] E. Çinlar, *Introduction to Stochastic Processes* (Prentice-Hall, Englewood Cliffs, 1975).

[3] J.W. Cohen, Basic problems of telephone traffic theory and the influence of repeated calls, Philips Telecomm. Rev. 18 (1957) 49–100.

[4] A. Erdélyi, W. Magnus, F. Oberhettinger and F.G. Tricomi, *Table of Integral Transforms*, Vol. 1 (McGraw-Hill, New York, 1954).

[5] G.I. Falin, A survey of retrial queues, Queueing Systems 7 (1990) 127–167.

[6] K. Farahmand, Single line queue with repeated demands, Queueing Systems 6 (1990) 223–228.

[7] G. Fayolle, A simple telephone exchange with delayed feedbacks, Traffic Anal. and Comput. Performance Evaluation (1986) 245.

[8] T. Hanschke, Bestimmung von Grenzwahrscheinlichkeiten bei Wartesschlangenmodellen mit Hilfe des Jacobi-Perron-Algorithmus, Berichte der math-stat Sektion im Forschungszentrum Graz 126 (1979).

[9] P. Hokstad, A supplementary variable technique applied to the $M/G/1$ queue, Scand. J. Stat. 2 (1975) 95–98.

[10] C.E.M. Pearce, Extended continued fractions, recurrence relations and two-dimensional Markov processes, Adv. Appl. Prob. 21 (1989) 357–375.

[11] B. Pourbabai, Analysis of a $G/M/K/0$ queueing system with heterogeneous servers and retrials, Int. J. Syst. Sci. 18 (1987) 985–992.
[12] T. Yang and J.G.C. Templeton, A survey on retrial queues, Queueing Systems 2 (1987) 203–233.
[13] T. Yang, A broad class of retrial queues and the associated generalized recursive technique, Ph.D. thesis, Univ. of Toronto (1990).