Queueing Systems 16 (1994) 83–96

The $M/G/1$ retrial queue with the server subject to starting failures

Tao Yang[1]

Department of Industrial Engineering, Technical University of Nova Scotia, Halifax, Nova Scotia, Canada B3J 2X4

Hui Li[2]

Department of Mathematics, Mount Saint Vincent University, Halifax, Nova Scotia, Canada B3M 2J6

Received 28 January, 1993; revised 9 July 1993

In this paper, we study a retrial queueing model with the server subject to starting failures. We first present the necessary and sufficient condition for the system to be stable and derive analytical results for the queue length distribution as well as some performance measures of the system in steady state. We show that the general stochastic decomposition law for $M/G/1$ vacation models also holds for the present system. Finally, we demonstrate that a few well known queueing models are special cases of the present model and discuss various interpretations of the stochastic decomposition law when applied to each of these special cases.

Keywords: Probability generating functions, retrial queues, steady state, server vacations, server setup times, stochastic decomposition.

1. Introduction

We study a single-server queueing system with the server subject to starting failures. New customers arrive according to a Poisson process with rate λ. We assume that there is no waiting space and therefore if an arriving customer finds the server busy or down, the customer makes a retrial at a later time. Returning customers behave independently of each other and are persistent in the sense that they keep making retrials until they receive their requested service, after which they have no further effects on the system. Successive inter-retrial times of any customer are independently, exponentially distributed with a common mean $1/\theta$. If the server is idle, an arriving (new or returning) customer must start or turn on the server, which takes zero time. If the server is started successfully (with a certain

[1] Partially supported by the Natural Sciences and Engineering Research Council of Canada, grant OGP0046415.
[2] Partially supported by internal research grant of Mount Saint Vincent University.

probability), the customer gets service immediately. Otherwise, the server undergoes "repair" immediately and the customer must leave and make a retrial at a later time. Successive service times and successive repairing times are independently, identically distributed with common distribution functions $B_1(\cdot)$ and $B_2(\cdot)$, respectively. We assume that the success probability is δ for a new customer who finds the server idle and sees no other customers in the system (the counterpart of a customer who starts a busy period in the standard $M/G/1$ system) and is α for *all other* new and returning customers.

The above model is a generalization of a few well known queues. For example, when $\delta = \alpha = 1$ it reduces to the $M/G/1$ retrial queue (Keilson et al. [12]) and when $\delta = 0$, $\alpha = 1$, and $\theta \to \infty$ it becomes the $M/G/1$ queueing system with setup times (Welch [25]) in which customers are served in random order. Queueing systems with customer retrials have attracted considerable attention in recent years. For related literature, interested readers may refer to two recent survey papers on the subject (Falin [8] and Yang and Templeton [26]). Here, we would like to point out, in particular, two related works: Kulkarni and Choi [16] and Choi et al. [2]. In the former, the authors derived analytical results for the $M/G/1$ retrial queue with server subject to repairs and breakdowns (*during* service or idle period) while in the latter, the authors studied the $M/G/1$ retrial queue with customer collisions. Queueing systems with setup times have also received considerable attention (Doshi [6], Lemoine [17, 18], Levy and Kleinrock [19], and Minh [21]). These models fall into the category of *queues with server vacations* on which a recent paper by Doshi [7] provides a comprehensive survey.

The paper is organized as follows. In the next section, we introduce necessary notations and assumptions, discuss the necessary and sufficient condition for system stability, and derive analytical results for the steady-state distribution of the number of customers in the system as well as formulas for some system performance measures. In section 3, we show that a general stochastic decomposition law for $M/G/1$ vacation systems also holds for the above system and discuss its interpretations when applied to various interesting special cases.

2. The analysis

We assume that both the service time distribution and the repair time distribution have finite first two moments, i.e.

$$\bar{x} \equiv \int_0^\infty x \, \mathrm{d}B_1(x), \quad \bar{x^2} \equiv \int_0^\infty x^2 \, \mathrm{d}B_1(x),$$

$$\bar{y} \equiv \int_0^\infty y \, \mathrm{d}B_2(y), \quad \bar{y^2} \equiv \int_0^\infty y^2 \, \mathrm{d}B_2(y).$$

We define, for $j = 1, 2, B_j^*(s) \equiv \int_0^\infty e^{-xs}\, dB_j(x)$ and $F_j(z) \equiv B_j^*(\lambda - \lambda z)$, and let $F(z) \equiv \alpha B_1^*(\lambda - \lambda z) + \bar{\alpha} z B_2^*(\lambda - \lambda z)$, where $\bar{\alpha} \equiv 1 - \alpha$. Furthermore, for convenience of presentation, we shall use the following notations: $\bar{\delta} \equiv 1 - \delta$, $\rho_1 \equiv \lambda \bar{x}$, and $\rho_2 \equiv \lambda \bar{y}$.

Let $N(t)$ be the number of customers in the process of returning at time t. Let $S(t)$ be a random variable such that

$$S(t) = \begin{cases} 0, & \text{if the server is idle at time } t, \\ 1, & \text{if the server is busy at time } t, \\ 2, & \text{if the server is down at time } t, \end{cases}$$

and $X(t)$ be the expended service time of the customer being served at time t if $S(t) = 1$ or the expended repair time at time t if $S(t) = 2$. $X(t) \equiv 0$ if $S(t) = 0$. Assuming that $S(0) = 0$, $N(0) = 0$, and $X(0) = 0$, we define ξ_n, $n = 0, 1, \ldots$, to be the instant at which the server starts its nth *idle* period, where $\xi_0 = 0$, and let $Q_n \equiv N(\xi_n^+)$. We note that ξ_n $(n \geq 1)$ is either a service completion epoch or a repair completion epoch and Q_n is the number of customers present in the system *immediately after* a service period or a repair period.

It can be easily seen that $\{Q_n;\, n \geq 0\}$ is a discrete-time Markov chain; $\{(Q_n, \xi_n);\, n \geq 0\}$ is a Markov renewal process (Çinlar [3]); and $\{(S(t), N(t), X(t)), t \geq 0\}$ is a continuous-time Markov process. Note that $\{(S(t), N(t), X(t)), t \geq 0\}$ is also a semi-regenerative process with $\{(Q_n, \xi_n);\, n \geq 0\}$ being its embedded Markov renewal process (Çinlar [3]).

We first study the necessary and sufficient condition for the system to be stable. We consider an arbitrary returning customer and assume that, at a particular visit, it finds the server idle and causes the server to undergo repair. Since its success probability is always α, the total number of *such* visits to the server before receiving service has a geometric distribution with mean $\bar{\alpha}/\alpha$. Recall that $\bar{y}$ is the mean repairing time. Thus, $\bar{\alpha}\bar{y}/\alpha$ is the expected total repairing time triggered by this particular customer. Since customers are persistent, the server must also spend on average $\bar{x}$ units of time serving the customer. Hence, in order to complete the service of *one returning customer*, the server must spend on average $\bar{x} + \bar{\alpha}\bar{y}/\alpha$ units of time during which $\lambda\bar{x} + \bar{\alpha}\lambda\bar{y}/\alpha$ *more* returning customers will arrive on average. Therefore, for the system to be stable, we must have $\lambda\bar{x} + \bar{\alpha}\lambda\bar{y}/\alpha < 1$.

The inequality $\lambda\bar{x} + \bar{\alpha}\lambda\bar{y}/\alpha < 1$ is also a sufficient condition for the system to be stable. To see this, we first prove that the embedded Markov chain $\{Q_n, n \geq 0\}$ is ergodic if $\lambda\bar{x} + \bar{\alpha}\lambda\bar{y}/\alpha < 1$. It is not difficult to see that $\{Q_n, n \geq 0\}$ is irreducible and aperiodic. To prove it is also positive recurrent, we shall use theorem 2 in Pakes [23] which states that an irreducible and aperiodic Markov chain $\{Q_n, n \geq 0\}$ is positive recurrent if $|\gamma_k| < \infty$ for all k and $\lim_{k\to\infty} \sup \gamma_k < 0$, where $\gamma_k = E(Q_{n+1} - Q_n | Q_n = k)$. In our case, we have $\gamma_0 = \delta\lambda\bar{x} + \bar{\delta}(1 + \lambda\bar{y})$ and for $k = 1, 2, \ldots$

$$\gamma_k = \frac{\alpha k\theta}{\lambda + k\theta}(k - 1 + \lambda\bar{x} - k) + \frac{\bar{\alpha}k\theta}{\lambda + k\theta}(k + \lambda\bar{y} - k) + \frac{\alpha\lambda}{\lambda + k\theta}(k + \lambda\bar{x} - k)$$
$$+ \frac{\bar{\alpha}\lambda}{\lambda + k\theta}(k + 1 + \lambda\bar{y} - k)$$
$$= -\alpha + \alpha\lambda\bar{x} + \bar{\alpha}\lambda\bar{y} + \frac{\lambda}{\lambda + k\theta}.$$

Clearly, if $\lambda\bar{x} + \bar{\alpha}\lambda\bar{y}/\alpha < 1$, then we have $|\gamma_k| < \infty$ for all k and $\lim_{k\to\infty} \sup \gamma_k < 0$. Therefore, the embedded Markov chain $\{Q_n;\, n \geq 0\}$ is ergodic. Since $\{(Q_n, \xi_n);\, n \geq 0\}$ is an embedded Markov renewal process of the semi-regenerative process $\{(S(t), N(t), X(t));\, t \geq 0\}$, it can be shown from the results in Çinlar [3, pp. 343–350], that the limiting probabilities of $\{(S(t), N(t), X(t));\, t \geq 0\}$ exist and are positive if $\lambda\bar{x} + \bar{\alpha}\lambda\bar{y}/\alpha < 1$ and $B_1(\cdot)$ and $B_2(\cdot)$ satisfy regular conditions (e.g., the existence of the first two moments or both $1 - B_1(x)$ and $1 - B_2(x)$ being Riemann integrable over $[0, \infty)$).

To summarize the above, we present

THEOREM 1

The inequality $\lambda\bar{x} + \bar{\alpha}\lambda\bar{y}/\alpha < 1$ is the necessary and sufficient condition for the system to be stable.

We will consider only the system in steady state and therefore we assume that $\lambda\bar{x} + \bar{\alpha}\lambda\bar{y}/\alpha < 1$ throughout this paper. For convenience, we denote $\rho \equiv \lambda\bar{x} + \bar{\alpha}\lambda\bar{y}/\alpha$. For the Markov process $\{(S(t), N(t), X(t));\, t \geq 0\}$, we define the unconditional probabilities $P_{0,k}(t) \equiv P\{S(t) = 0, N(t) = k\}$ for $t \geq 0$, $k = 0, 1, \ldots$, and the unconditional probability *densities* $P_{j,k}(x, t) \equiv P\{S(t) = j, N(t) = k, x \leq X(t) < x + \mathrm{d}x\}$ for $t \geq 0$, $j = 1, 2$, and $k = 0, 1, \ldots$. Following routine procedures (see, for example, Cox and Miller [5, pp. 262–266]), we obtain the equations that govern the dynamics of the system:

$$P'_{0,k}(t) = -(\lambda + k\theta)P_{0,k}(t) + \int_0^\infty r_1(x)P_{1,k}(x, t)\,\mathrm{d}x$$
$$+ \int_0^\infty r_2(x)P_{2,k}(x, t)\,\mathrm{d}x, \qquad k = 0, 1, \ldots, \qquad (1)$$

$$\frac{\partial}{\partial t}P_{1,k}(x,t) + \frac{\partial}{\partial x}P_{1,k}(x,t) = -[\lambda + r_1(x)]P_{1,k}(x,t) + \lambda P_{1,k-1}(x,t), \quad k = 0,1,\ldots, \tag{2}$$

$$\frac{\partial}{\partial t}P_{2,k}(x,t) + \frac{\partial}{\partial x}P_{2,k}(x,t) = -[\lambda + r_2(x)]P_{2,k}(x,t) + \lambda P_{2,k-1}(x,t), \quad k = 1,2,\ldots, \tag{3}$$

$$P_{1,0}(0,t) = \delta\lambda P_{0,0}(t) + \alpha\theta P_{0,1}(t), \tag{4}$$

$$P_{1,k}(0,t) = \alpha\lambda P_{0,k}(t) + \alpha(k+1)\theta P_{0,k+1}(t), \quad k = 1,2,\ldots, \tag{5}$$

$$P_{2,1}(0,t) = \bar{\delta}\lambda P_{0,0}(t) + \bar{\alpha}\theta P_{0,1}(t), \quad \text{and} \tag{6}$$

$$P_{2,k}(0,t) = \bar{\alpha}\lambda P_{0,k-1}(t) + \bar{\alpha}k\theta P_{0,k}(t), \quad k = 2,3,\ldots, \tag{7}$$

where $r_j(x) = B_j'(x)/[1 - B_j(x)]$, $j = 1,2$, and $P_{1,-1}(x,t) = P_{2,0}(x,t) = 0$ for any fixed x and t.

For the limiting probabilities $P_{0,k} \equiv \lim_{t\to\infty} P_{0,k}(t)$ and limiting densities $P_{j,k}(x) \equiv \lim_{t\to\infty} P_{j,k}(x,t)$, we define the generating functions (g.f.):

$$\phi_0(z) \equiv \sum_{k=0}^{\infty} P_{0,k}z^k, \quad \phi_j(x,z) \equiv \sum_{k=0}^{\infty} P_{j,k}(x)z^k, \quad \text{and} \quad \phi_j(z) \equiv \int_0^{\infty} \phi_j(x,z)\,\mathrm{d}x, \quad j = 1,2.$$

The main result is then given by

THEOREM 2

If $\rho < 1$, then

$$\phi_0(z) = \alpha(1-\rho)\frac{1 + (\delta - \alpha)c\varphi(0)\omega(z)}{\alpha + (\delta - \alpha)\varphi(0)[\alpha c\omega(1) - \rho_2]}\varphi(z), \tag{8}$$

$$\phi_1(z) = \frac{\alpha(1-\rho)[1 - F_1(z)]}{F(z) - z} \times \frac{\alpha\varphi(z) + (\delta - \alpha)\varphi(0)[\alpha c\omega(z)\varphi(z) - z(1 - F_2(z))/(1-z)]}{\alpha + (\delta - \alpha)\varphi(0)[\alpha c\omega(1) - \rho_2]}, \tag{9}$$

$$\phi_2(z) = \frac{\alpha(1-\rho)z[1 - F_2(z)]}{F(z) - z} \times \frac{\bar{\alpha}\varphi(z) + (\delta - \alpha)\varphi(0)[\bar{\alpha}c\omega(z)\varphi(z) - (F_1(z) - z)/(1-z)]}{\alpha + (\delta - \alpha)\varphi(0)[\alpha c\omega(1) - \rho_2]}, \tag{10}$$

where

$$c = \lambda/\theta, \varphi(z) = \exp\left\{ -c \int_1^z \frac{1 - F(u)}{u - F(u)}\, \mathrm{d}u \right\}$$

and

$$\omega(z) = \int_0^z \varphi^{-1}(u)\, \frac{F_1(u) - uF_2(u)}{u - F(u)}\, \mathrm{d}u.$$

Proof

Letting $t \to \infty$ in (1)–(7) and applying z-transform, we obtain

$$\lambda\phi_0(z) + \theta z\phi_0'(z) = \int_0^\infty \phi_1(x,z)r_1(x)\,\mathrm{d}x + \int_0^\infty \phi_2(x,z)r_2(x)\,\mathrm{d}x, \tag{11}$$

$$\frac{\partial\phi_j(x,z)}{\partial x} = -[\lambda(1-z) + r_j(x)]\phi_j(x,z), \quad j = 1,2, \tag{12}$$

$$\phi_1(0,z) = \alpha\lambda\phi_0(z) + (\delta - \alpha)\lambda P_{0,0} + \alpha\theta\phi_0'(z), \tag{13}$$

and

$$\phi_2(0,z) = \bar{\alpha}\lambda z\phi_0(z) - (\delta - \alpha)\lambda z P_{0,0} + \bar{\alpha}\theta z\phi_0'(z). \tag{14}$$

Solving (12), we have

$$\phi_j(x,z) = \phi_j(0,z)[1 - B_j(x)]\,\mathrm{e}^{-\lambda(1-z)x}, \quad j = 1,2. \tag{15}$$

Substituting the right hand sides of both (13) and (14) into (15), we obtain

$$\phi_1(x,z) = [\alpha\lambda\phi_0(z) + (\delta - \alpha)\lambda P_{0,0} + \alpha\theta\phi_0'(z)][1 - B_1(x)]\,\mathrm{e}^{-\lambda(1-z)x}, \tag{16}$$

$$\phi_2(x,z) = z[\bar{\alpha}\lambda\phi_0(z) - (\delta - \alpha)\lambda P_{0,0} + \bar{\alpha}\theta\phi_0'(z)][1 - B_2(x)]\,\mathrm{e}^{-\lambda(1-z)x}. \tag{17}$$

Applying the above to (11) and integrating, we have after rearrangement:

$$\phi_0'(z) + c\frac{1 - F(z)}{z - F(z)}\phi_0(z) = (\delta - \alpha)cP_{0,0}\frac{F_1(z) - zF_2(z)}{z - F(z)}. \tag{18}$$

Solving the above equation gives

$$\phi_0(z) = [\phi_0(1) + (\delta - \alpha)cP_{0,0}(\omega(z) - \omega(1))]\varphi(z), \tag{19}$$

where $\varphi(z)$ and $\omega(z)$ are defined in the theorem. To determine $\phi_0(1)$ and $P_{0,0}$, we set $z = 0$ in (19) and obtain $P_{0,0} = [\phi_0(1) - (\delta - \alpha)cP_{0,0}\omega(1)]\varphi(0)$ which in turn gives

$$P_{0,0} = \frac{\varphi(0)\phi_0(1)}{1 + (\delta - \alpha)c\omega(1)\varphi(0)}.$$

Replacing $P_{0,0}$ in (19) with the above, we obtain after algebraic manipulation:

$$\phi_0(z) = \phi_0(1)\frac{1 + (\delta - \alpha)c\varphi(0)\omega(z)}{1 + (\delta - \alpha)c\varphi(0)\omega(1)}\varphi(z). \tag{20}$$

Substituting (20) back into (16) and (17), we have

$$\phi_1(x,z) = \left[\alpha\lambda\frac{z-1}{z-F(z)}\phi_0(z) + \frac{(\delta-\alpha)\lambda\varphi(0)\phi_0(1)}{1+(\delta-\alpha)c\varphi(0)\omega(1)}\,\frac{z - zF_2(z)}{z-F(z)}\right] \times [(1 - B_1(x)]\,\mathrm{e}^{-\lambda(1-z)x}, \tag{21}$$

$$\phi_2(x,z) = z\left[\bar{\alpha}\lambda\frac{z-1}{z-F(z)}\phi_0(z) + \frac{(\delta-\alpha)\lambda\varphi(0)\phi_0(1)}{1+(\delta-\alpha)c\varphi(0)\omega(1)}\,\frac{F_1(z) - z}{z-F(z)}\right] \times [(1 - B_2(x)]\,\mathrm{e}^{-\lambda(1-z)x}. \tag{22}$$

Integrating the above from 0 to ∞ with respect to x, we obtain

$$\phi_1(z) = \int_0^\infty \phi_1(x,z)\,\mathrm{d}x = -\alpha\frac{1-F_1(z)}{z-F(z)}\phi_0(z) + \frac{(\delta-\alpha)\varphi(0)\phi_0(1)}{1+(\delta-\alpha)c\varphi(0)\omega(1)} \times \frac{z - zF_2(z)}{1-z}\,\frac{1-F_1(z)}{z-F(z)}, \tag{23}$$

$$\phi_2(z) = \int_0^\infty \phi_2(x,z)\,\mathrm{d}x = z\left[-\bar{\alpha}\frac{1-F_2(z)}{z-F(z)}\phi_0(z) + \frac{(\delta-\alpha)\varphi(0)\phi_0(1)}{1+(\delta-\alpha)c\varphi(0)\omega(1)} \times \frac{F_1(z) - z}{1-z}\,\frac{1-F_2(z)}{z-F(z)}\right.. \tag{24}$$

At this point, the only unknown is $\phi_0(1)$ which can be determined using the normalizing equation $\phi_0(1)+\phi_1(1)+\phi_2(1)=1$. Thus, setting $z=1$ in the above and applying l'Hôpital's rule whenever necessary, we obtain after rearragement

$$\begin{aligned}1 &= \phi_0(1)+\phi_1(1)+\phi_2(1)\\ &= \phi_0(1)+\frac{\alpha\rho_1}{\alpha(1-\rho)}\phi_0(1)+\frac{\bar{\alpha}\rho_2}{\alpha(1-\rho)}\phi_0(1)-\frac{\rho_1\rho_2}{\alpha(1-\rho)}\frac{(\delta-\alpha)\varphi(0)\phi_0(1)}{1+(\delta-\alpha)c\varphi(0)\omega(1)}\\ &\quad -\frac{(1-\rho_1)\rho_2}{\alpha(1-\rho)}\frac{(\delta-\alpha)\varphi(0)\phi_0(1)}{1+(\delta-\alpha)c\varphi(0)\omega(1)}.\end{aligned}$$

Solving the above equation for $\phi_0(1)$ yields

$$\phi_0(1)=\alpha(1-\rho)\frac{1+(\delta-\alpha)c\varphi(0)\omega(1)}{\alpha+(\delta-\alpha)\varphi(0)(\alpha c\omega(1)-\rho_2)}. \tag{25}$$

Substituting $\phi_0(1)$ into (20), (23), and (24), we obtain equations (8), (9), and (10). □

Remark

Define $P(z)=\phi_0(z)+z\phi_1(z)+\phi_2(z)$. Then, $P(z)$ can be considered as the probability generating function (p.g.f.) of the number of customers in the system (including the one in service if any) in steady state. As a direct consequence of theorem 2, we have

$$\begin{aligned}P(z) &= \frac{\alpha(1-\rho)(1-z)F_1(z)}{F(z)-z}\\ &\quad\times\frac{\alpha\phi(z)+(\delta-\alpha)\phi(0)[\alpha c\phi(z)\omega(z)-z(1-F_2(z))/(1-z)]}{\alpha+(\delta-\alpha)\phi(0)(\alpha c\omega(1)-\rho_2)}.\end{aligned} \tag{26}$$

We now derive formulas for some performance measures for the system in steady state. Let U be the server utilization (or the steady-state probability that the server is serving a customer), D be the steady state probability that the server is down, L be the average number of customers in the system in steady state, W be the average time a customer spends in the system in steady state, and R be the average number of retrials made by a customer. In the following, we present formulas only for U, D, and L since we have: $W=L/\lambda$ and $R=(W-\bar{x})\theta$ due to Little's formula. From theorem 2, we obtain:

$$U=\phi_1(1)=\rho_1, \tag{27}$$

$$D=\phi_2(1)=\rho_2\frac{\bar{\alpha}+(\delta-\alpha)\varphi(0)[\bar{\alpha}c\omega(1)-(1-\rho_1)]}{\alpha+(\delta-\alpha)\varphi(0)[\alpha c\omega(1)-\rho_2]}, \tag{28}$$

and

$$L = P'(1)$$

$$= \rho_1 + \frac{\alpha\lambda^2\overline{x^2} + \bar{\alpha}\lambda^2\overline{y^2} + 2\bar{\alpha}\rho_2 + 2c(\bar{\alpha} + \rho)}{2\alpha(1-\rho)}$$

$$- \frac{c(\delta - \alpha)\varphi(0)(\rho_2 + 1) + \rho_2 + \lambda^2\overline{y^2}/2}{\alpha + (\delta - \alpha)\varphi(0)[\alpha c\omega(1) - \rho_2]}. \tag{29}$$

3. Stochastic decomposition and special cases

Stochastic decomposition has been widely observed among $M/G/1$ type queues with server vacations (see, for example, Cooper [4], Levy and Yechiali [20], Fuhrmann [9], Heyman [11], Scholl and Kleinrock [24], and Doshi [6]). A key result in these analyses is that the number of customers in the system in steady state at a random point in time is distributed as the sum of two independent random variables, one of which is the number of customers in the corresponding standard queueing system in steady state at a random point in time. The other random variable may have different probabilistic interpretations in specific cases depending on how the vacations are scheduled. Stochastic decomposition has also been observed to hold for some $M/G/1$ retrial queues (Yang and Templeton [26]). A similar observation was made by Fuhrmann and Cooper [10] for a model which is technically the same as the standard $M/G/1$ retrial queue but arises in a different context (Neuts and Ramalhoto [22]).

For $M/G/1$ based vacation models, Fuhrmann and Cooper [10] reported a general stochastic decomposition law which states that the number of customers in *any* vacation system in steady state at a random point in time is distributed as the sum of two independent random variables: one being the number of customers in the corresponding standard $M/G/1$ system in steady state at a random point in time and the other being the number of customers in the vacation system at a random point in time *given* that the server is on vacation. Let $\pi(z)$ and $\chi(z)$ be the p.g.f.s of the first and the second random variables in the decomposition, respectively, and $\zeta(z)$ be the p.g.f. of the random variable being decomposed. Then, the mathematical version of the stochastic decomposition law is:

$$\zeta(z) = \pi(z)\chi(z). \tag{30}$$

We now verify that the decomposition law applies to the retrial model analyzed in the previous section. For the basic $M/G/1$ queue, we have:

$$\pi(z) = \frac{(1-\rho_1)(1-z)F_1(z)}{F_1(z) - z}. \tag{31}$$

To derive a formula for $\chi(z)$, we first define *vacation* in our context. We say that the server is on vacation if it is either under repair or idle. (Note that in retrial queues, there may be customers in the system even when the server is *idle*!) Under this definition, we have $\chi(z) = [\phi_0(z) + \phi_2(z)]/[\phi_0(1) + \phi_2(1)]$. Using the results of theorem 2, we obtain,

$$\chi(z) = \frac{\alpha(1-\rho)}{1-\rho_1}\frac{F_1(z)-z}{F(z)-z} \times \frac{\alpha\varphi(z) + (\delta-\alpha)\varphi(0)[\alpha c\varphi(z)\omega(z) - z(1-F_2(z))/(1-z)]}{\alpha + (\delta-\alpha)\varphi(0)(\alpha c\omega(1) - \rho_2)}. \tag{32}$$

From (26), we can see that $P(z) = \pi(z)\chi(z)$, which confirms that the decomposition law of Fuhrmann and Cooper [10] is also valid for this special vacation system. However, we must point out that if the idle periods were not considered as vacations, the decomposition law would not apply here (even if one use the $M/G/1$ retrial queue as the base system) due to interference between customer retrials and server vacations.

We now examine some special cases and discuss the corresponding implications of the stochastic decomposition law.

Case 1 ($\alpha = \delta = 1$). Here, the system reduces to the standard $M/G/1$ retrial queue (Keilson et al. [12]). From (26), the p.g.f. $P(z)$ becomes

$$P(z) = \frac{(1-\rho_1)(1-z)F_1(z)}{F_1(z)-z}\varphi(z), \tag{33}$$

where

$$\varphi(z) = \exp\left\{-c\int_1^z \frac{1-F_1(u)}{u-F_1(u)}\,du\right\}.$$

This is consistent with the results in Keilson et al. [12]. From theorem 2, we have $\phi_0(z) = (1-\rho_1)\varphi(z)$ and $\phi_2(z) = 0$ for this case and hence $\chi(z) = \varphi(z)$ is the p.g.f. of the number of customers in the system in steady state at a random point in time given that the server is idle (Fuhrmann and Cooper [10]).

Case 2 ($\alpha = 1, \delta = 0$, and $\theta \to \infty$). The system reduces to the $M/G/1$ queue with setup times (Welch [25]). Since the retrial rate θ tends to infinity, customers in the system will be always available for service and therefore when the server is idle there must be no customers in the system. Since $\delta = 0$ and $\alpha = 1$, the customer who starts a busy period always causes the server to warm up or to undergo repair with probability 1, after which all customers arrived in the busy period (including those, if any, arrived during the setup period) are served in random order. In this case, the generating function $P(z)$ becomes

$$P(z) = \frac{(1-\rho_1)(1-z)F_1(z)}{F_1(z)-z}\,\frac{1-zF_2(z)}{(1+\rho_2)(1-z)}, \tag{34}$$

which is consistent with the results in Welch [25]. According to the decomposition law, we have that

$$\chi(z) = \frac{1-zF_2(z)}{(1+\rho_2)(1-z)} \tag{35}$$

is the generating function of the number of customers present in the system in steady state at a random point in time given that the server is warming up. However, a more insightful interpretation follows from the observation that $\psi(z) \equiv zF_2(z)$ is the p.g.f. of the number of customers (including the one who starts the busy period) present in the system when the server returns from warming-up. From (35), we can rewrite $\chi(z)$ as

$$\chi(z) = \frac{1-\psi(z)}{\psi'(1)(1-z)}, \tag{36}$$

which implies that $\chi(z)$ is also the p.g.f. of the number of customers arrived during a time interval that is distributed as the equilibrium backward recurrence (expended) time of a warming-up period (Fuhrmann [9]).

Case 3 ($\alpha = \delta < 1$). This is the $M/G/1$ retrial queue in which the server is subject to starting failures. This model arose as a simplification of a situation where an equipment (the server) is shared by a large population of customers. Normally, the equipment is in the "off" state and must be turned on before it can be used by a customer, after which it is turned off again. There is a probability ($= \bar{\alpha}$) that the equipment may not be started successfully, in which case it undergoes repair immediately. For this system, $P(z)$ reduces to the following:

$$P(z) = \frac{\alpha(1-\rho)(1-z)F_1(z)}{F(z)-z}\varphi(z), \tag{37}$$

where $\varphi(z)$ is defined as in theorem 2. Decomposition can be carried out in two different ways depending on how we define vacation. If both repair periods and idle periods are considered as vacations, then the base system is the standard $M/G/1$ queue and $\pi(z)$ is given as in (31). The p.g.f. of the other random variable in the decomposition is then given by

$$\chi(z) = \frac{\alpha(1-\rho)}{1-\rho_1} \frac{F_1(z)-z}{F(z)-z} \varphi(z).$$

We now consider only the idle periods as vacations and the "repair" periods are treated as service sessions for customers who have triggered them. To appreciate this, we assume that when a (new or returning) customer arrives and finds the server idle, it requests either a *preparatory service session* (with probability $\bar{\alpha}$) or a *final service session* (with probability α). After each preparatory session, the customer leaves the server temporarily but makes retrials at later times for the rest of its service. After the final session, the customer leaves the system forever. The durations of preparatory sessions are independently, identically distributed with CDF $B_2(\cdot)$ and those of the final sessions are also independently, identically distributed but with CDF $B_1(\cdot)$. This is a queueing model with interrupted service times. It can be easily seen that this model is equivalent to the one with starting failures as far as the queue length distribution is concerned.

The total service requested by a customer consists of M preparatory sessions and one final session, where the random variable M is geometrically distributed with parameter $\bar{\alpha}$. Thus, the total service time has CDF $B(\cdot)$ whose Laplace transform is given as $\alpha B_1^*(s)/[1-\bar{\alpha}B_2^*(s)]$. Let $\eta(z)$ be the p.g.f. of the number of customers arriving during the total service time of a customer. Then, $\eta(z) = \alpha F_1/[1-\bar{\alpha}F_2(z)]$. In light of these, we can rewrite (37) as

$$P(z) = \frac{(1-\eta'(1))(1-z)\eta(z)}{\eta(z)-z} \varphi(z) \tag{38}$$

where the first fraction is the well known formula for a standard $M/G/1$ queue. From the above equation, we can conclude that, for the $M/G/1$ retrial queue with *interrupted service times*, the number of customers in the system at a random point in time is distributed as the sum of two independent random variables: one being the number of customers at a random point in time in a standard $M/G/1$ queue in which the service times are *not* interrupted and the other being the number of customers in the system at a random point in time *given* that the server is idle. However, for the $M/G/1$ retrial queue with *starting failures* (the original system), we can also conclude that the number of customers in the system can be considered as the sum of the number of customers in the standard $M/G/1$ queue (with service time distribution $B(\cdot)$ rather than $B_1(\cdot)$) and the number in the system *given* that the server is idle. Here, we note that for the *same* system stochastic decomposition can be interpreted differently depending on the definition of a vacation or the choice of the base system.

Case 4 ($\alpha = \delta$ and $B_1(\cdot) = B_2(\cdot)$). In this case, the system is in fact a "retrial" version of the $M/G/1$ queue with Bernoulli feedback (see, for example, Burke [1] and Kleinrock [13, 14]). Here, we assume instead that the setup period is a service period for the customer who triggers the server setup. After service completion, the customer may leave the system forever with probability α or may join the queue (the retrial group) again with probability $\bar{\alpha}$. From (26), we have the p.g.f. of the number of customers in the system:

$$P(z) = \frac{\alpha(1-\rho_1)(1-z)F_1(z)}{(\alpha+\bar{\alpha}z)F_1(z) - z}\varphi(z), \tag{39}$$

where

$$\varphi(z) = \exp\left\{-c\int_1^z \frac{1-(\alpha+\bar{\alpha}u)F_1(u)}{u-(\alpha+\bar{\alpha}u)F_1(u)}\,\mathrm{d}u\right\}.$$

It can be shown that the fraction on the right hand side is the p.g.f. of the $M/G/1$ queue with Bernoulli feedback (Kleinrock [15, p. 239]). Hence, we conclude that the number of customers in the $M/G/1$ retrial queue with Bernoulli feedback at a random point in time is distributed as the sum of two random variables: one being the number of customers at a random point in time in the corresponding $M/G/1$ queue with Bernoulli feedback and the other being the number of customers in the retrial system at a random point in time given that the server is idle.

In summary, we have investigated a retrial queueing model with the server subject to starting failures. We have presented the necessary and sufficient condition for the system to be stable and derived analytical results for the queue length distribution as well as some performance measures of a system in steady state. We have shown that the general stochastic decomposition law for $M/G/1$ vacation models also holds for the system. Finally, we have demonstrated that a few well known queueing models are special cases of the present model and discussed various interpretations of the stochastic decomposition law when applied to each of these special cases.

Acknowledgements

We would like to thank the anonymous referees, the associate editor, and the editor-in-chief for their helpful comments. In particular, we are grateful for the referees' suggestions of using Pakes' theorem and a probabilistic argument rather than lengthy mathematical derivations in the proof of the necessary and sufficient condition for system stability.

References

[1] P.J. Burke, Proof of a conjecture on the interarrival time distribution in an $M/M/1$ queue with feedback, IEEE Trans. Commun. 24 (1976) 175–178.

[2] B.D. Choi, Y.W. Shin and W.C. Ahn, Retrial queues with collision arising from unslotted CSMA/CD protocol, Queueing Syst. 11 (1992) 335–356.

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

[4] R.B. Cooper, Queues served in cyclic order: waiting times, Bell Syst. Tech. J. 49 (1970) 399–413.

[5] D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes* (Methuen, London, 1965).

[6] B.T. Doshi, A note on stochastic decomposition in a $GI/G/1$ queue with vacations or set-up times, J. Appl. Prob. 22 (1985) 419–428.

[7] B.T. Doshi, Queueing systems with vacations – A survey, Queueing Syst. 1 (1986) 29–66.

[8] G. Falin, Retrial queues, Queueing Syst. 7 (1990) 127–168.

[9] S.W. Fuhrmann, A note on the $M/G/1$ queue with server vacations, Oper. Res. 32 (1984) 1368–1373.

[10] S.W. Fuhrmann and R.B. Cooper, Stochastic decomposition in the $M/G/1$ queue with generalized vacations, Oper. Res. 33 (1985) 1117–1129.

[11] D.P. Heyman, Optimal operating policies for $M/G/1$ queueing systems, Oper. Res. 16 (1968) 362–382.

[12] J. Keilson, J. Cozzolino and H. Young, A service system with unfilled requests repeated, Oper. Res. 16 (1968) 1126–1137.

[13] L. Kleinrock, Analysis of a time-shared processor, Nav. Res. Log. Quart. 11 (1964) 59–73.

[14] L. Kleinrock, Time-shared systems: a theoretical treatment, J. ACM 14 (1967) 242–261.

[15] L. Kleinrock, *Queueing Systems*, Vol. I: *Theory* (Wiley, New York, 1975).

[16] V.G. Kulkarni and B.D. Choi, Retrial queue with server subject to breakdowns and repairs, Queueing Syst. 7 (1990) 191–208.

[17] A.J. Lemoine, Limit theorems for generalized single server queue, Adv. Appl. Prob. 6 (1974) 159–174.

[18] A.J. Lemoine, Limit theorems for generalized single server queue: the exceptional system, SIAM J. Appl. Math. 28 (1975) 596–606.

[19] H. Levy and L. Kleinrock, A queue with starter and a queue with vacations: delay analysis by decomposition, Oper. Res 34 (1986) 426–436.

[20] Y. Levy and U. Yechiali, Utilization of idle time in an $M/G/1$ queueing system, Manag. Sci. 22 (1975) 202–211.

[21] D. Le Minh, Analysis of the exceptional queueing system by the use of regenerative processes and analytical methods, Math. Oper. Res. 5 (1980) 147–159.

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

[23] A.G. Pakes, Some conditions for ergodicity and recurrence of Markov chains, Oper. Res. 17 (1969) 1058–1061.

[24] M. Scholl and L. Kleinrock, On the $M/G/1$ queue with rest period and certain service-independent queueing disciplines, Oper. Res. 31 (1983) 705–719.

[25] P.D. Welch, On a generalized $M/G/1$ queueing process in which the first customer of each busy period receives exceptional service, Oper. Res. 12 (1964) 736–752.

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