Queueing Systems 38, 363–380, 2001

Reliability Analysis of the Retrial Queue with Server Breakdowns and Repairs *

JINTING WANG ** and JINHUA CAO
Institute of Applied Mathematics, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing 100080, P.R. China

QUANLIN LI
Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P.R. China

Received 2 January 2000; Revised 1 March 2001

Abstract. Retrial queues have been widely used to model many problems arising in telephone switching systems, telecommunication networks, computer networks and computer systems, etc. It is of basic importance to study reliability of retrial queues with server breakdowns and repairs because of limited ability of repairs and heavy influence of the breakdowns on the performance measure of the system. However, so far the repairable retrial queues are analyzed only by queueing theory. In this paper we give a detailed analysis for reliability of retrial queues. By using the supplementary variables method, we obtain the explicit expressions of some main reliability indexes such as the availability, failure frequency and reliability function of the server. In addition, some special queues, for instance, the repairable $M/G/1$ queue and repairable retrial queue can be derived from our results. These results may be generalized to the repairable multi-server retrial models.

Keywords: reliability, retrial queues, supplementary variable method

1. Introduction

During the last two decades considerable attention has paid to the analysis of queuing systems with repeated calls (or retrial queues, queues with returning customers, etc.) see, for example, the surveys by Yang and Templeton [25], Falin [12], Kulkarni and Liang [19], and the book by Falin and Templeton [14]. Retrial queueing systems are characterized by the feature that arrivals who find the server busy join the retrial group (orbit) to try again for their requests in random order and at random intervals. These queueing models have been used to model many problems in telephone switching systems,telecommunication networks, computer networks, computer and communication systems. Some details on applications research in the field can be found in the proceed-

* Project supported by the National Natural Sciences Foundation of China (No. 19871093).
** Current address: Department of Mathematics, Northern Jiaotong University, Beijing, 100044, P.R. China.

ings of successive International Teletrafic Congresses and in the review papers or book mentioned above.

In this paper, we study a single-server system with the server subject to breakdowns and repairs. Such system with repairable server has been studied as queueing models by many authors such as Avi-Itzhak and Naor [5], Thiruvengdan [24], Mitrany and Avi-Itzhak [20], Neuts and Lucantoni [21]. Meanwhile, such systems can also be regarded as reliability models. Cao and Cheng [6] considered the model in [5] from the viewpoint of reliability for the first time. Although it is natural in the real world, there are only a few works take into consideration retrial phenomena involving the unreliability of the server. For related literature, interested readers may refer to [1,2,4,18,26] where a single-server queue with unreliable server was considered.

The first result on $M/G/1$ retrial queues is due to Keilson et al. [16] who used the method of supplementary variables. They obtained the joint distribution of the channel state and the queue length in steady state. Later, Aleksandrov [3] and Falin [9] independently of Keilson et al. [16] considered the case of general service times using different methods. Some of its variations have been studied by a number of authors including, for example, Falin [9,11], Kulkarni [17], Choi and Park [7], Neuts and Ramalhoto [22], Farahmand [15], and Yang and Li [26,27]. The primary focus of this paper is to give a detailed analysis for reliability of such retrial queues. By using the supplementary variables method, we obtain the explicit expressions of some main reliability indexes such as the availability, failure frequency and reliability function of the server. These results may be generalized to the repairable multi-server retrial models.

The rest of the paper is organized as follows. In the next section, we give a relative formal description of the queueing model and some definitions and conventions. In section 3, we investigate the steady-state analytic formulas for our model by using the supplementary variable method. In section 4 we obtain steady-state solutions to the model, and derive the queue length and waiting time of the model. In the final section 5 we consider the reliability quantities of the server, and the availability, failure frequency, reliability function are obtained.

2. Description of the model

We consider a single-server queueing system in which new customers (primary calls) arrival in a Poisson process with rate λ. We assume that there is no waiting space and therefore if an arriving customer finds the server idle, the customer obtains service immediately and leaves the system after service completion. Otherwise, if the server is found busy or down, the customer makes a retrial at a later time and then the arriving customer becomes a source of repeated calls (a customer in retrial group). The pool of sources of repeated calls may be viewed as a sort of queue with infinite capacity. 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 $1/\theta$.

We assume that the server fails at an exponential rate α, i.e., the server fails after an exponential amount time with mean $1/\alpha$. When the server fails, it is repaired immediately and the time required to repair it is a random variable Y with general distribution $G(\cdot)$ and p.d.f. $g(\cdot)$ defined by

$$g(y) = \beta(y) \exp\left\{-\int_0^y \beta(t)\,\mathrm{d}t\right\},$$

and the service times of customers are independent, identical distributed with common general distribution functions $B(\cdot)$ and p.d.f. $b(\cdot)$ defined by

$$b(x) = \mu(x) \exp\left\{-\int_0^x \mu(t)\,\mathrm{d}t\right\},$$

where $\beta(y)$ and $\mu(x)$ are the repair completion rate and the service completion rate as follows:

$$\beta(y) = \frac{g(y)}{1 - G(y)}, \qquad \mu(x) = \frac{b(x)}{1 - B(x)}.$$

Throughout this paper we use $*$ representing the Laplace transform, for example,

$$f^*(s) = \int_0^\infty f(t) \exp\{-st\}\,\mathrm{d}t, \quad \mathrm{Re}(s) > 0,$$

and $\overline{f^*}(s) \triangleq 1 - f^*(s)$. We assume that both the service time distribution and the repair time distribution have finite first two moments: $\beta_k = (-1)^k b^{*(k)}(0)$ and $\gamma_k = (-1)^k g^{*(k)}(0)$, $k = 1, 2$. Denote by $\rho = \lambda\beta_1$ the system load due to primary calls.

Assume that the service time for a customer is cumulative, and after repair the server is as good as new. The input flow of primary calls, intervals between repetitions, service times, failure times and repair times are mutually independent.

3. The steady-state equations

We first study the condition for system stability. The following theorem provides the necessary and sufficient condition for the system to be stable.

Theorem 3.1. The inequality $\rho(1 + \alpha\gamma_1) < 1$ is the necessary and sufficient condition for the system to be stable.

Proof. To prove the necessity of the condition, we define $\widetilde{\chi}_n$ to be the generalized service time of the nth customer, that is, the length of time since the nth customer begins to be served until the service is completed, where $\widetilde{\chi}_n$ includes some possible down times of the server due to server failures during the service period of the nth customer. It is obvious that $\widetilde{\chi}_n$ is independent of n, $n = 1, 2, \ldots$.

By assumption, repeated calls have no effects on $\widetilde{\chi}_n$. Hence, some results obtained in [6], where the classical $M/G/1$ queueing system with repairable server was studied, can be used here. In order to obtain the distribution function of $\widetilde{\chi}_n$, $n = 1, 2, \ldots$, define

$$\begin{aligned} \widetilde{B}_n^{[l]}(t) = \Pr\{&\widetilde{\chi}_n \leqslant t \text{ and server just fails } l \text{ times during the interval since the} \\ &n\text{th customer begins to be served until the service is completed}\}, \\ n \geqslant 1,\ l \geqslant 0,& \end{aligned} \tag{3.1}$$

for $t \geqslant 0$. Then, it can be shown in [6] that the generalized successive service times $\{\widetilde{\chi}_n\}$ are identically distributed, independent random variables with distribution function

$$\widetilde{B}(t) \triangleq \widetilde{B}_n(t) = \Pr\bigl\{\widetilde{\chi}_n \leqslant t\bigr\} = \sum_{l=0}^{\infty} \int_0^t G^{(l)}(t-u)\mathrm{e}^{-\alpha u} \frac{(\alpha u)^l}{l!}\,\mathrm{d}B(u), \tag{3.2}$$

which is independent of n. Its Laplace–Stieltjes transform is

$$\widetilde{b}^*(s) = \int_0^{\infty} \mathrm{e}^{-st}\,\mathrm{d}\widetilde{B}(t) = b^*\bigl(s + \alpha - \alpha g^*(s)\bigr), \quad \mathrm{Re}(s) > 0, \tag{3.3}$$

and its expected value is given by

$$\mathrm{E}\widetilde{\chi}_n = -\frac{\mathrm{d}\widetilde{b}^*(s)}{\mathrm{d}s}\bigg|_{s=0} = \beta_1(1 + \alpha\gamma_1). \tag{3.4}$$

From (3.4), we can see that in order to complete the service of one customer, the server must spend on average $\beta_1(1+\alpha\gamma_1)$ units of time during which $\lambda\beta_1(1+\alpha\gamma_1)$ more customers will arrive on average. Therefore, for the system to be stable, we must have $\rho(1 + \alpha\gamma_1) < 1$.

The inequality $\rho(1 + \alpha\gamma_1) < 1$ is also a sufficient condition for the system to be stable. To prove this statement, we first prove the embedded Markov chain $\{Q_n,\ n \geqslant 0\}$ is ergodic if $\rho(1 + \alpha\gamma_1) < 1$. It is readily to see that $\{Q_n,\ n \geqslant 0\}$ is irreducible and aperiodic. It remains to be proved that $\{Q_n,\ n \geqslant 0\}$ is positive recurrent. We use [23, theorem 2] which states that an irreducible and aperiodic Markov chain $\{Q_n, n \geqslant 0\}$ is positive recurrent if $|\Psi_k| < \infty$ for all k and $\lim_{k\to\infty} \sup \Psi_k < 0$, where $\Psi_k = E(Q_{n+1} - Q_n \mid Q_n = k)$. In our case, for $k = 0, 1, 2, \ldots$

$$\begin{aligned} \Psi_k &= \frac{k\theta}{\lambda + k\theta}\bigl[k - 1 + \lambda\beta_1(1 + \alpha\gamma_1) - k\bigr] + \frac{\lambda}{\lambda + k\theta}\bigl[k + \lambda\beta_1(1 + \alpha\gamma_1) - k\bigr] \\ &= \lambda\beta_1(1 + \alpha\gamma_1) - \frac{k\theta}{\lambda + k\theta}. \end{aligned}$$

Obviously, if $\rho(1+\alpha\gamma_1) < 1$, then we have $|\Psi_k| < \infty$ for all k and $\lim_{k\to\infty} \sup \Psi_k < 0$. Therefore, the embedded Markov chain $\{Q_n;\ n \geqslant 0\}$ is ergodic. It can be shown from the results in [8] that if $\rho(1+\alpha\gamma_1) < 1$ and $G(\cdot)$ and $B(\cdot)$ satisfy regular conditions then the system is stable. $\square$

Remark 1. A lemma in [6] states that if the $\widetilde{\chi}_n$ is considered as the "service time" of the nth customer, then the $M/G/1$ queueing system with repairable server can be investigated as a classical $M/G/1$ queueing system where the input process is a homogeneous Poisson process with rate λ and the successive service times $\{\widetilde{\chi}_n\}$ are identically distributed, independent random variables with distribution function $\widetilde{B}(t)$. To differ from the ordinary $M/G/1$ queueing system, the notation $M/\widetilde{G}/1$ is used to define the new queueing system in that paper.

Remark 2. Denote $\widetilde{\pi}$ by the busy period of the $M/\widetilde{G}/1$ queueing system, and define

$$\widetilde{\Pi}(t) = \Pr\{\widetilde{\pi} \leqslant t\}, \quad t \geqslant 0,$$
$$\widetilde{\pi}(s) = \int_0^{\infty} \mathrm{e}^{-st}\,\mathrm{d}\widetilde{\Pi}(t), \quad R(s) \geqslant 0,$$

it is shown in [6] that $\widetilde{\pi}(s)$ is the root with smallest absolute value of the equation

$$z = \widetilde{b}^*(s + \lambda - \lambda z). \tag{3.5}$$

We shall consider the system in steady state which exits if and only if $\rho(1 + \alpha\gamma_1) < 1$, so the condition $\rho(1 + \alpha\gamma_1) < 1$ is assumed to hold from now on.

As in the ordinary $M/G/1$ queues, the stochastic process $\{N(t);\ t \geqslant 0\}$, where $N(t)$ is the number of customers in the system (both in orbit and in service) at time t, is not Markovian. Therefore, we introduce random variables $X(t)$, $Y(t)$, where $X(t)$ is the elapsed service time of the customer in service at time t, and $Y(t)$ is the elapsed repair time of the server at time t. And define the following state probabilities:

(1) $P_{Wi1}(t, x)\,\mathrm{d}x$ is the joint probability that at time t there are i customers in the retrial group, the server is up and a customer is being served with elapsed service time between x and $x + \mathrm{d}x$ ($i \geqslant 0$).

(2) $P_{Ri1}(t, x, y)\,\mathrm{d}y$ is the joint probability that at time t there are i customers in the retrial group, the elapsed service time for the customer under service is equal to x, and the server is being repaired with the elapsed repair time between y and $y + \mathrm{d}y$ ($i \geqslant 0$).

(3) $P_{Ii0}(t)$ is the probability that the server is idle at time t, and there are i customers in the retrial group.

Thus the stochastic process $\{(N(t), X(t), Y(t));\ t \geqslant 0\}$ is Markovian with state space $\{(I, i, 0),\ (W, i, 1, x),\ (R, i, 1, x, y) \mid 0 \leqslant i < +\infty,\ 0 \leqslant x, y < +\infty\}$. By considering transitions of the process between time t and $t + \Delta t$ and letting $\Delta t \to 0$, we derive the system of forward equations for $i = 0, 1, 2, \ldots$:

$$\left[\frac{\mathrm{d}}{\mathrm{d}t} + \lambda + i\theta\right] P_{Ii0}(t) = \int_0^{\infty} \mu(x) P_{Wi1}(t, x)\,\mathrm{d}x, \tag{3.6a}$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\mu(x)+\lambda+\alpha\right]P_{Wi1}(t,x)=\int_0^{\infty}\beta(y)P_{Ri1}(t,x,y)\,\mathrm{d}y+\lambda P_{W,i-1,1}(t,x), \tag{3.6b}$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+\lambda+\beta(y)\right]P_{Ri1}(t,x,y)=\lambda P_{R,i-1,1}(t,x,y). \tag{3.6c}$$

Equations (3.6a)–(3.6c) are to be solved under the following boundary conditions:

$$P_{Ri1}(t,x,0)=\alpha P_{Wi1}(t,x), \tag{3.6d}$$
$$P_{Wi1}(t,0)=\lambda P_{Ii0}(t)+(i+1)\theta P_{I,i+1,0}(t) \tag{3.6e}$$

together with the normalizing equation

$$\sum_{i=0}^{+\infty}\left\{P_{Ii0}(t)+\int_0^{+\infty}P_{Wi1}(t,x)\,\mathrm{d}x+\int_0^{+\infty}\int_0^{+\infty}P_{Ri1}(t,x,y)\,\mathrm{d}x\,\mathrm{d}y\right\}=1,$$

and an initial condition

$$P_{I00}(0)=1,$$

where $P_{W,-1,0}(t,x)\triangleq 0$, $P_{R,-1,0}(t,x,y)\triangleq 0$ for any fixed t, x and y.

Since we interested in the steady-state behaviour of the system, let us assume that the condition for the system to be stable is satisfied and set $P_{Ii0}\triangleq\lim_{t\to\infty}P_{Ii0}(t)$, $P_{Wi1}(x)\triangleq\lim_{t\to\infty}P_{Wi1}(t,x)$, $P_{Ri1}(x,y)\triangleq\lim_{t\to\infty}P_{Ri1}(t,x,y)$, then we can derive from the forward equations above that

$$(\lambda+i\theta)P_{Ii0}=\int_0^{\infty}\mu(x)P_{Wi1}(x)\,\mathrm{d}x, \tag{3.7a}$$
$$\left[\frac{\mathrm{d}}{\mathrm{d}x}+\mu(x)+\lambda+\alpha\right]P_{Wi1}(x)=\int_0^{\infty}\beta(y)P_{Ri1}(x,y)\,\mathrm{d}y+\lambda P_{W,i-1,1}(x), \tag{3.7b}$$
$$\left[\frac{\partial}{\partial y}+\lambda+\beta(y)\right]P_{Ri1}(x,y)=\lambda P_{R,i-1,1}(x,y), \tag{3.7c}$$
$$P_{Ri1}(x,0)=\alpha P_{Wi1}(x), \tag{3.7d}$$
$$P_{Wi1}(0)=\lambda P_{Ii0}+(i+1)\theta P_{I,i+1,0}. \tag{3.7e}$$

The normalizing equation becomes

$$\sum_{i=0}^{+\infty}\left\{P_{Ii0}+\int_0^{+\infty}P_{Wi1}(x)\,\mathrm{d}x+\int_0^{+\infty}\int_0^{+\infty}P_{Ri1}(x,y)\,\mathrm{d}x\,\mathrm{d}y\right\}=1. \tag{3.7f}$$

4. The model solution

To solve the system of equations (3.7a)–(3.7e), we define the following generating functions:

$$Q_{I0}(z)=\sum_{i=0}^{\infty}P_{Ii0}z^i,\qquad Q_{W1}(z,x)=\sum_{i=0}^{\infty}P_{Wi1}(x)z^i,$$
$$Q_{R1}(z,x,y)=\sum_{i=0}^{\infty}P_{Ri1}(x,y)z^i,\quad |z|\leqslant 1.$$

Multiplying equations (3.7a)–(3.7e) by z^i and summing over i, we obtain the following basic equations after some algebraic manipulations:

$$\lambda Q_{I0}(z)+z\theta Q'_{I0}(z)=\int_0^{\infty}\mu(x)Q_{W1}(z,x)\,\mathrm{d}x, \tag{4.1a}$$
$$\left[\frac{\partial}{\partial x}+\mu(x)+\lambda+\alpha\right]Q_{W1}(z,x)=\int_0^{\infty}\beta(y)Q_{R1}(x,y)\,\mathrm{d}y+\lambda zQ_{W,1}(z,x), \tag{4.1b}$$
$$\left[\frac{\partial}{\partial y}+\lambda+\beta(y)\right]Q_{R1}(z,x,y)=\lambda zQ_{R,1}(z,x,y), \tag{4.1c}$$
$$Q_{R1}(z,x,0)=\alpha Q_{W1}(z,x), \tag{4.1d}$$
$$Q_{W1}(z,0)=\lambda Q_{I0}(z)+\theta Q'_{I,0}(z). \tag{4.1e}$$

Solving the differential equation (4.1c), we obtain

$$Q_{R1}(z,x,y)=Q_{R1}(z,x,0)\mathrm{e}^{-\lambda(1-z)y}\overline{G}(y)=\alpha Q_{W1}(z,x)\mathrm{e}^{-\lambda(1-z)y}\overline{G}(y). \tag{4.2}$$

Combining equations (4.1b) and (4.2) we get

$$\begin{aligned}\left[\frac{\partial}{\partial x}+\mu(x)+\lambda+\alpha\right]&Q_{W1}(z,x)\\ &=\int_0^{\infty}\beta(y)Q_{R1}(z,x,0)\mathrm{e}^{-\lambda(1-z)y}\overline{G}(y)\,\mathrm{d}y+\lambda zQ_{W1}(z,x)\\ &=\alpha Q_{W1}(z,x)g^*\big(\lambda(1-z)\big)+\lambda zQ_{W1}(z,x).\end{aligned}$$

Solving above differential equation, we can obtain

$$Q_{W1}(z,x)=Q_{W1}(z,0)\exp\big\{-\big[\lambda(1-z)+\alpha\overline{g^*}\big(\lambda(1-z)\big)\big]x\big\}\overline{B}(x). \tag{4.3}$$

Equation (4.1e) can be rewritten as

$$Q_{W1}(z,0)=\theta Q'_{I0}(z)+\lambda Q_{I0}(z),$$

together with equations (4.1a) and (4.3), we obtain

$$\theta z Q'_{I0}(z) + \lambda Q_{I0}(z) = \int_0^\infty \mu(x)\big(\theta Q'_{I0}(z) + \lambda Q_{I0}(z)\big) \times \exp\big\{-\big[\lambda(1-z) + \alpha\overline{g^*}\big(\lambda(1-z)\big)\big]x\big\}\overline{B}(x)\,\mathrm{d}x = \big[\theta Q'_{I0}(z) + \lambda Q_{I0}(z)\big]b^*\big(\lambda(1-z) + \alpha\overline{g^*}\big(\lambda(1-z)\big)\big),$$

or

$$\theta\big[z - b^*\big(\lambda(1-z) + \alpha\overline{g^*}(\lambda - \lambda z)\big)\big]Q'_{I0}(z) = \lambda b^*\big(\lambda(1-z) + \alpha\overline{g^*}(\lambda - \lambda z)\big)Q_{I0}(z) - \lambda Q_{I0}(z).$$

This is a first order differential equation with nonconstant coefficients and can be solved by standard methods to obtain the following general solution:

$$Q_{I0}(z) = C\exp\left\{-\frac{\lambda}{\theta}\int_z^1 \frac{b^*[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))] - 1}{u - b^*[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))]}\,\mathrm{d}u\right\}. \tag{4.4}$$

Therefore

$$Q_{W1}(z,0) = \theta Q'_{I0}(z) + \lambda Q_{I0}(z) = \frac{\lambda(z-1)}{z - b^*(\lambda(1-z) + \alpha\overline{g^*}(\lambda - \lambda z))}Q_{I0}(z). \tag{4.5}$$

Let $z \to 1$ in (4.5), we obtain by the L'Hospital's rule that

$$Q_{W1}(1,0) = \frac{\lambda}{1 - \rho(1 + \alpha\gamma_1)}Q_{I0}(1),$$

where $\rho = \lambda\beta_1$. This gives

$$Q_{W1}(1,x) = \frac{\lambda}{1 - \rho(1 + \alpha\gamma_1)}Q_{I0}(1)\overline{B}(x),$$

$$Q_{R1}(1,x,y) = \frac{\alpha\lambda}{1 - \rho(1 + \alpha\gamma_1)}Q_{I0}(1)\overline{B}(x)\overline{G}(y).$$

From the total probability

$$Q_{I0}(1) + \int_0^\infty Q_{W1}(1,x)\,\mathrm{d}x + \int_0^\infty\int_0^\infty Q_{R1}(1,x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$$

we obtain $C = 1 - \rho(1 + \alpha\gamma_1)$ as the probability that the server is idle.

Thus we summarize our results in the following theorem.

Theorem 4.1. In the steady state, the joint distribution of the server state and queue length has partial generating function:

$$Q_{I0}(z) = \big[1 - \rho(1 + \alpha\gamma_1)\big] \times \exp\left\{-\frac{\lambda}{\theta}\int_z^1 \frac{b[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))] - 1}{u - b[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))]}\,\mathrm{d}u\right\}, \tag{4.6}$$

$$
\begin{aligned}
Q_{W1}(z,x) &= \frac{\lambda(z-1)}{z - b^*(\lambda(1-z) + \alpha\overline{g^*}(\lambda-\lambda z))} \\
&\quad \times \exp\bigl\{-\bigl[\lambda(1-z) + \alpha\overline{g^*}\bigl(\lambda(1-z)\bigr)\bigr]x\bigr\}\overline{B}(x)Q_{I0}(z), \qquad (4.7)\\
Q_{R1}(z,x,y) &= \frac{\alpha\lambda(z-1)}{z - b^*(\lambda(1-z) + \alpha\overline{g^*}(\lambda-\lambda z))} \\
&\quad \times \exp\bigl\{-\bigl[\lambda(1-z) + \alpha\overline{g^*}\bigl(\lambda(1-z)\bigr)\bigr]x\bigr\}\overline{B}(x) \\
&\quad \times \exp\bigl\{-\lambda(1-z)y\bigr\}\overline{G}(y)Q_{I0}(z). \qquad (4.8)
\end{aligned}
$$

Corollary 4.2. If the system is in steady state, then

(1) the probability that the server is idle is $I = 1 - \rho(1 + \alpha\gamma_1)$;

(2) the probability that the server is busy is $B = \rho$;

(3) the probability that the server under repair is $R = \rho\alpha\gamma_1$.

Proof. Note that

$$
\begin{aligned}
I &= \lim_{z\to 1} Q_{I0}(z), \qquad B = \lim_{z\to 1}\int_0^\infty Q_{W1}(z,x)\,\mathrm{d}x,\\
R &= \lim_{z\to 1}\int_0^\infty\int_0^\infty Q_{R1}(z,x,y)\,\mathrm{d}x\,\mathrm{d}y,
\end{aligned}
$$

the stated formulas follow by direct calculations. □

Corollary 4.3. The distribution of the number of repeated calls $q_n = P\{N_1(t) = n\}$ has generating function:

$$
\begin{aligned}
p(z) &= Q_{I0}(z) + Q_{W1}(z) + Q_{R1}(z) \\
&= \bigl[1 - \rho(1+\alpha\gamma_1)\bigr]\frac{1-z}{b^*[\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))] - z} \\
&\quad \times \exp\biggl\{-\frac{\lambda}{\theta}\int_z^1 \frac{b^*[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))] - 1}{u - b^*[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))]}\,\mathrm{d}u\biggr\}. \qquad (4.9)
\end{aligned}
$$

In particular, the mean queue length in orbit is given by

$$
\mathrm{E}N(t) = \frac{\lambda^2}{1-\rho(1+\alpha\gamma_1)}\biggl[\frac{\alpha\beta_1\gamma_2 + (1+\alpha\gamma_1)^2\beta_2}{2} + \frac{(1+\alpha\gamma_1)\beta_1}{\theta}\biggr]. \qquad (4.10)
$$

Proof. Integrating (4.7) with respect to x and using well known formula

$$
\int_0^\infty \mathrm{e}^{-sx}\bigl(1 - B(x)\bigr)\,\mathrm{d}x = \frac{1-b^*(s)}{s},
$$

we have

$$Q_{W1}(z) = \int_0^\infty Q_{W1}(z,x)\,\mathrm{d}x = \frac{\lambda - \lambda z}{b^*[\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))] - z} \times \frac{1 - b^*[\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))]}{\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))} Q_{I0}(z). \quad (4.11)$$

Similar analysis yields

$$Q_{R1}(z) = \int_0^\infty \int_0^\infty Q_{R1}(z,x,y)\,\mathrm{d}x\,\mathrm{d}y = \alpha \frac{\overline{b^*}[\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))]}{b^*(\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))) - z} \times \frac{\overline{g^*}(\lambda - \lambda z)}{\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))} Q_{I0}(z). \quad (4.12)$$

With the help of generating functions $Q_{I0}(z),\ Q_{W1}(z),\ Q_{R1}(z)$, we get the distribution of the number of repeated calls $q_n = P\{N_1(t) = n\}$ has generating function:

$$\begin{aligned} p(z) &= Q_{I0}(z) + Q_{W1}(z) + Q_{R1}(z) \\ &= \left[1 - \rho(1+\alpha\gamma_1)\right] \frac{1-z}{b^*[\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))] - z} \\ &\quad \times \exp\left\{-\frac{\lambda}{\theta}\int_z^1 \frac{b^*[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))] - 1}{u - b^*[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))]}\,\mathrm{d}u\right\}. \end{aligned}$$

By direct calculation we can obtain (4.10). □

Corollary 4.4. The distribution of the number of customers in the system $Q_n = P\{K(t) = n\}$ has generating function:

$$\begin{aligned} Q(z) &= \left[1 - \rho(1+\alpha\gamma_1)\right] \frac{(1-z)b^*[\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))]}{b^*[\lambda(1-z) + \alpha\overline{g^*}(\lambda(1-z))] - z} \\ &\quad \times \exp\left\{-\frac{\lambda}{\theta}\int_z^1 \frac{b^*[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))] - 1}{u - b^*[\lambda(1-u) + \alpha\overline{g^*}(\lambda(1-u))]}\,\mathrm{d}u\right\}. \end{aligned} \quad (4.13)$$

In particular, the mean queue length $\mathrm{E}K(t)$ is given by

$$\mathrm{E}K(t) = \rho(1+\alpha\gamma_1) + \frac{\lambda^2}{1 - \rho(1+\alpha\gamma_1)}\left[\frac{\alpha\beta_1\gamma_2 + (1+\alpha\gamma_1)^2\beta_2}{2} + \frac{(1+\alpha\gamma_1)\beta_1}{\theta}\right].$$

Proof. This is readily obtained by considering the following equation:

$$Q(z) = Q_{I0}(z) + zQ_{W1}(z) + zQ_{R1}(z),$$

and the mean queue length can be readily derived from (4.13). □

Remark 3. When $\alpha = 0$, our model becomes the $M/G/1$ retrial queue with reliable server. In this case, (4.6) and (4.7) reduce to

$$Q_{I0}(z) = (1-\rho)\exp\left\{-\frac{\lambda}{\theta}\int_z^1 \frac{b(\lambda(1-u))-1}{u-b(\lambda(1-u))}\,\mathrm{d}u\right\},$$
$$Q_{W1}(z,x) = \frac{\lambda(z-1)}{z-b^*(\lambda(1-z))}\exp\bigl\{-(\lambda-\lambda z)x\bigr\}\overline{B}(x)Q_{I0}(z),$$

which agree with [12, equations (10), (11)].

The generating function of the queue length in the orbit is $Q_{I0}(z)+Q_{W1}(z)$ which equals

$$(1-\rho)\frac{1-z}{b^*(\lambda(1-z))-z}\exp\left\{-\frac{\lambda}{\theta}\int_z^1 \frac{b^*(\lambda(1-u))-1}{u-b^*(\lambda(1-u))}\,\mathrm{d}u\right\},$$

the generating function of the queue length in the system is $Q_{I0}(z)+zQ_{W1}(z)$ which equals

$$(1-\rho)\frac{(1-z)b^*(\lambda(1-z))}{b^*(\lambda(1-z))-z}\exp\left\{-\frac{\lambda}{\theta}\int_z^1 \frac{b^*(\lambda(1-u))-1}{u-b^*(\lambda(1-u))}\,\mathrm{d}u\right\},$$

and the mean queue length of the queue is

$$\mathrm{E}N(t) = \frac{\lambda^2}{1-\rho}\left(\frac{\beta_2}{2}+\frac{\beta_1}{\theta}\right), \qquad \mathrm{E}K(t) = \rho + \frac{\lambda^2}{1-\rho}\left(\frac{\beta_2}{2}+\frac{\beta_1}{\theta}\right).$$

These results are consistent with known results in [14].

Remark 4. As were used in [6], the method of Markov renewal process and Takács' results can be used to obtain these results presented above.

Let us now consider the waiting time process of the system. The mean waiting time of W in the steady state can be easily obtained with the help of Little's formula:

$$W = \frac{\mathrm{E}N(t)}{\lambda} = \frac{\lambda}{1-\rho(1+\alpha\gamma_1)}\left[\frac{\alpha\beta_1\gamma_2+(1+\alpha\gamma_1)^2\beta_2}{2}+\frac{(1+\alpha\gamma_1)\beta_1}{\theta}\right].$$

However, the analysis of waiting time in retrial queues is much more difficult than the number of customers in the system due to the fact that retrial queues are queues with random overtaking.

Suppose that at the moment of departure of some customer (considered as the 0th customer) there are $n \geqslant 1$ other customers in the system. Tag one of them and denote by T_n its waiting time. Let

$$\widetilde{\pi}^{(k)}(s,y) \triangleq \mathrm{E}\bigl\{\mathrm{e}^{-s\widetilde{L}^{(k)}}y^{\tilde{I}^{(k)}}\bigr\}$$

be the Laplace transform of the joint distribution of the length of the k-busy period $\widetilde{L}^{(k)}$ and the number of the demands which were served during this k-busy period, $\tilde{I}^{(k)}$, in our model. It is clear that it is different from which defined in [13], because the k-busy period in our model may includes the repair time. By using known results in [6,13], we can obtain the following theorem.

Theorem 4.5. The Laplace transform of the waiting time of the tagged customer is given by the formula:

$$\begin{aligned}\mathrm{E}e^{-sT_n} = \int_1^{\widetilde{\pi}(s)} & \frac{u^{n-1}}{b^*(s+\lambda-\lambda u+\alpha\overline{g^*}(s+\lambda(1-u)))-u} \\ & \times \exp\left\{\int_u^1 \frac{\theta+s+\lambda\overline{b^*}(s+\lambda-\lambda v+\alpha\overline{g^*}(s+\lambda(1-v)))}{\theta[b^*(s+\lambda-\lambda v+\alpha\overline{g^*}(s+\lambda(1-v)))-v]}\,\mathrm{d}v\right\}\mathrm{d}u, \quad (4.14)\end{aligned}$$

where $\widetilde{\pi}(s)$ is the root with smallest absolute value of the equation given by

$$z = b^*\big(s+\lambda-\lambda z+\alpha\overline{g^*}(s+\lambda-\lambda z)\big). \quad (4.15)$$

Remark 5. We can see that $\widetilde{\pi}(s)$ is the Laplace transform of the length of a busy period in the corresponding $M/\widetilde{G}/1$ queue defined in [6].

5. Reliability indexes of the server

We now consider some reliability quantities of the server in this section. Let

$$A(t) = P\{\text{the service station is up at time } t\}$$

which is defined as the pointwise availability of the server, and define the steady-state availability of the server as $A = \lim_{t\to\infty} A(t)$.

Theorem 5.1. The steady-state availability of the server is $A = 1 - \rho\alpha\gamma_1$.

Proof. This is readily obtained by considering the following equation:

$$A = \sum_{i=0}^{\infty} P_{Ii0} + \sum_{i=0}^{\infty}\int_0^{\infty} P_{Wi1}(x)\,\mathrm{d}x = \lim_{z\to 1}\left[Q_{I0}(z) + \int_0^{\infty} Q_{W1}(z,x)\,\mathrm{d}x\right],$$

together with (4.6) and (4.11). □

Theorem 5.2. The steady-state failure frequency of the server is $W_f = \alpha\rho$.

Proof. Since the steady-state failure frequency of the service station is

$$W_f = \sum_{i=0}^{\infty}\int_0^{+\infty} \alpha P_{Wi1}(x)\,\mathrm{d}x,$$

we get

$$W_f = \lim_{z\to 1} \int_0^{+\infty} \alpha Q_{W1}(z,x)\,\mathrm{d}x = \alpha\rho. \qquad \square$$

Denote by τ the time to the first failure of the server, then the reliability function of the server is

$$R(t) = P(\tau > t).$$

Theorem 5.3. The Laplace transform of $R(t)$ is given by

$$\begin{aligned} R^*(s) = {} & \frac{1}{s+\alpha} + \frac{\alpha}{s+\alpha} \int_1^w \frac{1}{\theta[b^*(s+\alpha+\lambda(1-y)) - y](1-y)} \\ & \times \exp\left\{ \frac{1}{\theta} \int_y^1 \frac{s+\lambda-\lambda b^*(s+\alpha+\lambda(1-x))}{b^*(s+\alpha+\lambda(1-x)) - x}\,\mathrm{d}x \right\} \mathrm{d}y, \end{aligned} \tag{5.1}$$

where ω is the root of the equation $z = b^*(s+\alpha+\lambda(1-z))$ inside $|z| = 1$, $\mathrm{Re}(s) > 0$.

Proof. In order to find the reliability of the server, letting the failure states of the server be absorbing states, then we obtain a new system. In the new system, we use the same notations as in the previous section, then we can get the following set of equations:

$$\left[\frac{\mathrm{d}}{\mathrm{d}t} + \lambda + i\theta\right] P_{Ii0}(t) = \int_0^\infty \mu(x) P_{Wi1}(t,x)\,\mathrm{d}x, \tag{5.2a}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial x} + \lambda + \alpha + \mu(x)\right] P_{Wi1}(t,x) = \lambda P_{W,i-1,1}(t,x), \tag{5.2b}$$

$$P_{Wi1}(t,0) = (i+1)\theta P_{I,i+1,0}(t) + \lambda P_{Ii0}(t), \tag{5.2c}$$

with the initial condition: $P_{I00}(0) = 1$.

By taking Laplace transforms of these equations, we obtain

$$sP^*_{Ii0}(s) - 1 = -(\lambda + i\theta)P^*_{Ii0}(s) + \int_0^\infty \mu(x) P^*_{Wi1}(s,x)\,\mathrm{d}x, \tag{5.3a}$$

$$sP^*_{Wi1}(s,x) + \frac{\partial P^*_{Wi1}(s,x)}{\partial x} = -\big(\lambda+\alpha+\mu(x)\big)P^*_{Wi1}(s,x) + \lambda P^*_{W,i-1,1}(s,x), \tag{5.3b}$$

$$P^*_{Wi1}(s,0) = (i+1)\theta P^*_{I,i+1,0}(s) + \lambda P^*_{Ii0}(s). \tag{5.3c}$$

Define the following generating functions

$$Q^*_{I0}(s,z) = \sum_{i=0}^{\infty} P^*_{Ii0}(s) z^i, \qquad Q^*_{W1}(s,z,x) = \sum_{i=0}^{\infty} P^*_{Wi1}(s,x) z^i.$$

Multiplying equations (5.3a)–(5.3c) by z^i and summing over i, we obtain the following basic equations after some algebraic manipulations:

$$(s+\lambda)Q_{I0}^*(s,z)-\sum_{i=0}^{\infty}z^i=-z\theta\frac{\partial Q_{I0}^*(s,z)}{\partial z}+\int_0^{\infty}\mu(x)Q_{W1}^*(s,z,x)\,\mathrm{d}x,\quad(5.4\mathrm{a})$$

$$sQ_{W1}^*(s,z,x)+\frac{\partial Q_{W1}^*(s,z,x)}{\partial x}=-\big(\lambda+\alpha+\mu(x)\big)Q_{W1}^*(s,z,x)+\lambda zQ_{W1}^*(s,z,x),\quad(5.4\mathrm{b})$$

$$Q_{W1}^*(s,z,0)=\theta\frac{\partial Q_{I0}^*(s,z)}{\partial z}+\lambda Q_{I0}^*(s,z).\quad(5.4\mathrm{c})$$

From (5.4b) and (5.4c) we obtain

$$\begin{aligned}Q_{W1}^*(s,z,x)&=Q_{W1}^*(s,z,0)\exp\big\{-(s+\lambda+\alpha-\lambda z)x\big\}\overline{B}(x)\\&=\left[\lambda Q_{I0}^*(s,z)+\theta\frac{\partial Q_{I0}^*(s,z)}{\partial z}\right]\exp\big\{-(s+\lambda+\alpha-\lambda z)x\big\}\overline{B}(x).\end{aligned}\quad(5.5)$$

Substituting (5.5) into (5.4a) yields

$$\begin{aligned}&(s+\lambda)Q_{I0}^*(s,z)+z\theta\frac{\partial Q_{I0}^*(s,z)}{\partial z}\\&\quad=\sum_{i=0}^{\infty}z^i+\left[\lambda Q_{I0}^*(s,z)+\theta\frac{\partial Q_{I0}^*(s,z)}{\partial z}\right]b^*\big(s+\alpha+\lambda(1-z)\big),\end{aligned}$$

this gives

$$\begin{aligned}&\theta\big[b^*\big(s+\alpha+\lambda(1-z)-z\big)\big]\frac{\partial Q_{I0}^*(s,z)}{\partial z}\\&\quad=\big[s+\lambda-\lambda b^*\big(s+\alpha+\lambda(1-z)\big)\big]Q_{I0}^*(s,z)-\sum_{i=0}^{\infty}z^i.\end{aligned}\quad(5.6)$$

Consider the coefficient $f(z)=b^*(s+\alpha+\lambda(1-z))-z$. It is easy to see that

$$\begin{aligned}&f(0)=b^*(s+\alpha+\lambda)\geqslant 0,\\&f(1)=b^*(s+\alpha)-1<0,\\&f''(z)=\lambda^2b^{*\prime\prime}\big(s+\alpha+\lambda(1-z)\big)\geqslant 0,\quad\text{i.e., function } f(x) \text{ is convex.}\end{aligned}$$

Thus $f(x)$ has exactly one root ω in the interval [0, 1]. If $0\leqslant z<\omega$, then the coefficient $b^*(s+\alpha+\lambda(1-z))-z$ is strictly positive. Hence the general solution of (5.6) on the interval $0\leqslant z<\omega$ is

$$Q_{I0}^*(s,z)=e(s,z)\left\{C_1-\int_0^z\frac{1}{\theta[b^*(s+\alpha+\lambda(1-y))-y](1-y)e(s,y)}\,\mathrm{d}y\right\},\quad(5.7)$$

where

$$e(s,z) = \exp\left\{\frac{1}{\theta}\int_0^z \frac{s+\lambda-\lambda b^*(s+\alpha+\lambda(1-y))}{b^*(s+\alpha+\lambda(1-y))-y}\,\mathrm{d}y\right\}, \tag{5.8}$$

for $0 \leqslant z < \omega$.

When $y \to \omega$, the function under the integral sign on the right-hand side of (5.7) is an infinite of an order $1/(\omega - y)$. Then the integral

$$\int_0^\omega \frac{s+\lambda-\lambda b^*(s+\alpha+\lambda(1-y))}{b^*(s+\alpha+\lambda(1-y))-y}\,\mathrm{d}y$$

is devergent and so the function $e(s,z) \to \infty$ as $z \to \omega$. On the other hand, $Q^*_{I0}(s,z) < \infty$. This allows us to determine the constant C_1:

$$C_1 = \int_0^\omega \frac{1}{\theta[b^*(s+\alpha+\lambda(1-y))-y](1-y)e(s,y)}\,\mathrm{d}y$$

and therefore (5.7) is reduced to

$$Q^*_{I0}(s,z) = e(s,z)\int_z^\omega \frac{1}{\theta[b^*(s+\alpha+\lambda(1-y))-y](1-y)e(s,y)}\,\mathrm{d}y, \tag{5.9}$$

for $0 \leqslant z < \omega$.

Similarly, we can consider the case $\omega < z < 1$. When $\omega < z \leqslant 1$ the coefficient $b^*(s+\alpha+\lambda(1-z)) - z$ is strictly negative and so

$$Q^*_{I0}(s,z) = e_1(s,z)\left\{C_2 - \int_1^z \frac{1}{\theta[b^*(s+\alpha+\lambda(1-y))-y](1-y)e_1(s,y)}\,\mathrm{d}y\right\}, \tag{5.10}$$

where

$$e_1(s,z) = \exp\left\{\frac{1}{\theta}\int_1^z \frac{s+\lambda-\lambda b^*(s+\alpha+\lambda(1-y))}{b^*(s+\alpha+\lambda(1-y))-y}\,\mathrm{d}y\right\}, \tag{5.11}$$

and $\omega < z \leqslant 1$. In the same manner as above, it follows from (5.10) and (5.11) that

$$Q^*_{I0}(s,z) = e_1(s,z)\int_\omega^z \frac{1}{\theta[b^*(s+\alpha+\lambda(1-y))-y](1-y)e_1(s,y)}\,\mathrm{d}y, \tag{5.12}$$

where $\omega < z \leqslant 1$.

After some algebra, (5.9) and (5.12) can be reduced to the joint formula

$$Q^*_{I0}(s,z) = \int_z^\omega \frac{1}{\theta[b^*(s+\alpha+\lambda(1-y))-y](1-y)} \times \exp\left\{\frac{1}{\theta}\int_y^z \frac{s+\lambda-\lambda b^*(s+\alpha+\lambda(1-x))}{b^*(s+\alpha+\lambda(1-x))-x}\,\mathrm{d}x\right\}\mathrm{d}y,$$

if $z \neq \omega$.

For $z = \omega$, we have directly from (5.6) that

$$Q_{I0}^*(s, \omega) = \frac{1}{(1-\omega)(s+\lambda-\lambda\omega)}.$$

So, we can obtain

$$\begin{aligned}
Q_{W1}^*(s, z) &= \int_0^\infty Q_{W1}^*(s, z, x)\,\mathrm{d}x \\
&= \Bigg[\frac{1}{z - b^*(s+\alpha+\lambda(1-z))} \\
&\quad + \left(\lambda + \frac{\lambda b^*(s+\alpha+\lambda(1-z)) - \lambda - s}{z - b^*(s+\alpha+\lambda(1-z))}\right) Q_{I0}^*(s, z)\Bigg] \\
&\quad \times \frac{\overline{b^*}(s+\alpha+\lambda-\lambda z)}{s+\alpha+\lambda-\lambda z} \\
&= \left[\frac{1}{z - b^*(s+\alpha+\lambda(1-z))} + \frac{(\lambda z - \lambda - s) Q_{I0}^*(s, z)}{z - b^*(s+\alpha+\lambda(1-z))}\right] \\
&\quad \times \frac{\overline{b^*}(s+\alpha+\lambda-\lambda z)}{s+\alpha+\lambda-\lambda z}.
\end{aligned} \tag{5.13}$$

Hence, we have

$$\begin{aligned}
R^*(s) &= Q_{I0}^*(s, 1) + Q_{W1}^*(s, 1) \\
&= \left\{1 + \left(-\frac{s}{(1 - b^*(s+\alpha))}\right) \frac{\overline{b^*}(s+\alpha)}{s+\alpha}\right\} Q_{I0}^*(s, 1) + \frac{1}{s+\alpha} \\
&= \frac{\alpha}{s+\alpha} Q_{I0}^*(s, 1) + \frac{1}{s+\alpha}.
\end{aligned}$$

Upon substitution, we obtain formula (5.1). □

From theorem 5.3 we obtain:

Corollary 5.4. The mean time to the first failure (MTTFF) of the server is given by

$$\begin{aligned}
\mathrm{MTTFF} &= \frac{1}{\alpha} + \int_1^w \frac{1}{\theta[b^*(\alpha+\lambda(1-y)) - y](1-y)} \\
&\quad \times \exp\left\{\frac{1}{\theta}\int_y^1 \frac{\lambda - \lambda b^*(\alpha+\lambda(1-x))}{b^*(\alpha+\lambda(1-x)) - x}\,\mathrm{d}x\right\} \mathrm{d}y,
\end{aligned} \tag{5.14}$$

where ω is the root of the equation $z = b^*(s+\alpha+\lambda(1-z))$ inside $|z| = 1$.

Proof. From (5.1) and the following equation

$$\mathrm{MTTFF} = \int_0^\infty R(t)\,\mathrm{d}t = R^*(s)|_{s=0},$$

we obtain (5.14). □

Acknowledgements

The authors would like to thank the referee for valuable suggestions and comments that help to improve the presentation of this paper.

References

[1] A. Aissani, Unreliable queueing with repeated orders, Microelectron. Reliab. 33 (1993) 2093–2106.
[2] A. Aissani, A retrial queue with redundancy and unreliable server, Queueing Systems 17 (1995) 431–449.
[3] A.M. Aleksandrov, A queueing system with repeated orders, Engrg. Cybernet. Rev. 12(3) (1974) 1–4.
[4] J.R. Artalejo, New results in retrial queueing systems with breakdown of the servers, Statistica Neerlandica 48 (1) (1994) 23–36.
[5] B. Avi-Itzhak and P. Naor, Some queueing problems with the service station subject to breakdown, Oper. Res. 10 (1962) 303–320.
[6] J. Cao and K. Cheng, Analysis of $M/G/1$ queueing system with repairable service station, Acta Math. Appl. Sinica 5 (1982) 113–127.
[7] B.D. Choi and K.K. Park, The $M/G/1$ retrial queue for Bernoulli schedule, Queueing Systems 7 (1990) 219–228.
[8] E. Çinlar, *Introduction to Stochastic Processes* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
[9] G.I. Falin, Multi-phase servicing in a single-channel system for automation of experiments with repeated calls, in: *Problems of Automation of Scientific Investigations in Radio Engineering and Electronics* (USSR Academy of Science, Moscow, 1975).
[10] G.I. Falin, Aggregate arrival of customers in one-line system with repeated calls, Ukr. Math. J. 28 (1976) 437–440.
[11] G.I. Falin, On the waiting-time process in a single-line queue with repeated calls, J. Appl. Probab. 23 (1986) 185–192.
[12] G.I. Falin, A survey of retrial queues, Queueing Systems 7 (1990) 127–167.
[13] G.I. Falin and C. Fricker, On the virtual waiting time in an $M/G/1$ retrial queue, J. Appl. Probab. 28 (1991) 446–460.
[14] G.I. Falin and J.G.C. Templeton, *Retrial Queues* (Chapman & Hall, London, 1997).
[15] K. Farahmand, Single line queue with repeated demands, Queueing Systems 6 (1990) 223–228.
[16] J. Keilson, J. Cozzolino and H. Young, A service system with unfilled requests repeated, Oper. Res. 16 (1968) 1126–1137.
[17] V.G. Kulkarni, On queueing systems with retrials, J. Appl. Probab. 20 (1983) 380–389.
[18] V.G. Kulkarni and B.D. Choi, Retrial queues with server subject to breakdowns and repairs, Queueing Systems 7 (1990) 191–208.
[19] V.G. Kulkarni and H.M. Liang, Retrial queues revisited, in: *Frontiers in Queueing*, ed. J.H. Dshalalow (CRC Press, Boca Raton, FL, 1997).
[20] I.L. Mitrany and B. Avi-Itzhak, A many server queue with service interruptions, Oper. Res. 16 (1968) 628–638.
[21] M.F. Neuts and D.M. Lucantoni, A Markovian queue with N servers subject to breakdowns and repairs, Managm. Sci. 25 (1979) 849–861.
[22] 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.
[23] A.G. Pakes, Some conditions for ergodicity and recurrence of Markov chains, Oper. Res. 17 (1969) 1058–1061.
[24] K. Thiruvenydan, Queueing with breakdown, Oper. Res. 11 (1963) 62–71.
[25] T. Yang and J.G.C. Templeton, A survey on retrial queues, Queueing Systems 2 (1987) 201–233.

[26] T. Yang and H. Li, The $M/G/1$ retrial queue with the server subject to starting failures, Queueing Systems 16 (1994) 83–96.
[27] T. Yang and H. Li, On the steady-state queue size distribution of the discrete-time $Geo/G/1$ queue with repeated customers, Queueing Systems 21 (1995) 199–215.