RAIRO-Oper. Res. 54 (2020) 471–488
https://doi.org/10.1051/ro/2019008

RAIRO Operations Research
www.rairo-ro.org

AN M/G/1 RETRIAL QUEUE WITH SINGLE WORKING VACATION UNDER BERNOULLI SCHEDULE

Tao Li[1,*], Liyuan Zhang[2] and Shan Gao[3]

Abstract. In this paper, an M/G/1 retrial queue with general retrial times and single working vacation is considered. We assume that the customers who find the server busy are queued in the orbit in accordance with a first-come-first-served (FCFS) discipline and only the customer at the head of the queue is allowed access to the server. During the normal period, if the orbit queue is not empty at a service completion instant, the server begins a working vacation with specified probability q $(0 \leq q \leq 1)$, and with probability $1 - q$, he waits for serving the next customer. During the working vacation period, customers can be served at a lower service rate. We first present the necessary and sufficient condition for the system to be stable. Using the supplementary variable method, we deal with the generating functions of the server state and the number of customers in the orbit. Various interesting performance measures are also derived. Finally, some numerical examples and cost optimization analysis are presented.

Mathematics Subject Classification. 60k25, 90B22.

Received February 15, 2016. Accepted January 1, 2019.

1. Introduction

Retrial queueing systems are described by the feature that the arriving customers who find the server busy join the retrial orbit to try again for their requests. Review of retrial queue literature could be found in Falin and Templeton [7], Artalejo and Corral [11], Falin [6] and Artalejo [10]. The concept of Bernoulli vacation schedule was first introduced by Keilson and Servi [8]. If the queue is empty after a service completion then the server becomes inactive, *i.e.*, a vacation period begins. If the queue is not empty then another service begins with specified probability p or a vacation period begins with probability $1 - p$. Many researchers have paid attention to the retrial queueing models with vacation under Bernoulli schedule. Kumar and Arivudainambi [2] studied an M/G/1 retrial queue with Bernoulli vacation and general retrial times, Zhou [29] considered a similar model but with setup time. Kumar *et al.* [3] generalized the model of [2] to an M^X/G/1 queue. Some authors like Choudhury and Ke [5], Wang [12] and Wu and Lian [13] also discussed a retrial queue with vacation under Bernoulli schedule.

On the basis of ordinary vacation, Servi and Finn [15] first introduced the concept of working vacation, where the server provides service at a lower speed during the vacation period rather than stopping the service completely. Since queueing models with working vacation can be applied in manufacturing systems, service systems and communication systems, working vacation has become an important aspect. Using the matrix-analytic method, Tian *et al.* [20] analyzed an M/M/1 queue with single working vacation, Li and Tian [9] studied a GI/M/1 queue. Using the method of supplementary variable, Zhang and

Keywords. Retrial, working vacation, Bernoulli, embedded Markov chain; supplementary variable method.

[1] School of Mathematics and Statistics, Shandong University of Technology, Zibo 255049, China.

[2] Business School, Shandong University of Technology, Zibo 255049, China.

[3] School of Mathematics and Statistics, Fuyang Normal College, Fuyang 236037, China.

*Corresponding author: liltaot@126.com

Article published by EDP Sciences

Hou [18] discussed an M/G/1 queue with single working vacation. Readers can also refer to Chae *et al.* [14], Selvaraju and Goswami [19] and Gao and Liu [25].

Recently, the retrial queueing systems with working vacation have been investigated extensively. Do [27] first studied an M/M/1 retrial queue with working vacations which is motivated by the performance analysis of a Media Access Control function in wireless networks. Using the matrix-analytic method, Tao *et al.* [16] considered an M/M/1 retrial queue with collisions and working vacation interruption under N-policy, Upadhyaya [26] analyzed a discrete-time Geo^X/Geo/1 retrial queue with working vacations. Using the method of supplementary variable, Aissani *et al.* [1] and Jailaxmi *et al.* [28] both generalized the model of [27] to an M/G/1 queue with constant retrial policy, and Arivudainambi *et al.* [4] analyzed an M/G/1 queue with general retrial time policy. Gao *et al.* [24] discussed an M/G/1 retrial queue with general retrial times and working vacation interruption, the discrete-time Geo^X/G/1 queue was investigated by Gao and Wang [23]. Although there have been some works about retrial queue with working vacation, the server only begins a working vacation when the system becomes empty. To the authors' best knowledge, there is no research work investigating retrial queue with working vacation under Bernoulli schedule. This motivates us to deal with such a queueing model in this paper. Obviously, let parameters in this paper take proper values, many working vacation queues will be the special cases of the model we consider.

This paper is organized as follows. In Section 2, we give a brief description of the model. The stability condition is obtained by the matrix-analytic method in Section 3. In Section 4, we deal with the joint distribution of the server state and the number of customers in the orbit. Some performance measures of this model are discussed in Section 5. Section 6 presents numerical examples and cost optimization analysis. Finally, Section 7 concludes the paper.

2. System model

In this paper, we consider an M/G/1 retrial queue with general retrial times, and the server takes a working vacation under Bernoulli schedule. The detailed description of this model is given as follows:

(1) Customers arrive according to a Poisson process with rate λ, and there is no waiting space in front of the server. If the customer finds the server free when he arrives, he will begin his service immediately. If the server is busy, the arriving customer will join the retrial orbit.
(2) We assume that only the customer at the head of the orbit queue is allowed access to the server, and the retrial time R follows an arbitrary distribution with distribution function $R(x)$.
(3) In a regular busy period, the normal service time S_b is assumed to be generally distributed with distribution function $G_b(x)$ and first two moments μ_1, μ_2. During a working vacation period, the lower service time S_v follows an arbitrary distribution with distribution function $G_v(x)$ and first two moments η_1, η_2.
(4) At a service completion instant in the normal period, if the orbit is not empty, the server takes a working vacation with probability q $(0 \le q \le 1)$, and with probability $\overline{q}$ $(\overline{q} = 1 - q)$, he waits for serving the next customer. If the orbit is empty, the server always takes a working vacation. Vacation time V follows an exponential distribution with parameter θ. At the end of each vacation, the server starts a new busy period if there are customers in the system. Otherwise, the server stays idle and will serve the new arrival by normal service rate.

It is assumed that $R(0) = 0, R(\infty) = 1, G_b(0) = 0, G_b(\infty) = 1, G_v(0) = 0, G_v(\infty) = 1$, and $R(x)$, $G_b(x)$, $G_v(x)$ are continuous at $x = 0$. The functions $\alpha(x)$, $\mu(x)$ and $\eta(x)$ are the conditional completion rates for retrial time, normal service time and lower service time, respectively, *i.e.*,

$$\alpha(x)\,\mathrm{d}x = \frac{\mathrm{d}R(x)}{1 - R(x)}, \quad \mu(x)\,\mathrm{d}x = \frac{\mathrm{d}G_b(x)}{1 - G_b(x)}, \quad \eta(x)\,\mathrm{d}x = \frac{\mathrm{d}G_v(x)}{1 - G_v(x)}.$$

Further, we assume that all the random variables defined above are independent. Throughout the rest of the paper, for a distribution function $F(x)$, we define $\overline{F}(x) = 1 - F(x)$ to be the tail of $F(x)$, $\widetilde{F}(s) = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}F(x)$ to be the Laplace–Stieltjes transform (LST) of $F(x)$ and $\overline{F}^*(s) = \int_0^\infty \mathrm{e}^{-sx}\overline{F}(x)\,\mathrm{d}x$ to be the Laplace transform of $\overline{F}(x)$. Clearly, we can obtain that $\overline{F}^*(s) = \frac{1-\widetilde{F}(s)}{s}$.

Let $N(t)$ represent the number of customers in the retrial orbit at time t, and $I(t)$ denote the server state, defined as follows

$$I(t) = \begin{cases} 0, & \text{the server is in a working vacation period at time } t \text{ and the server is free,} \\ 1, & \text{the server is in a working vacation period at time } t \text{ and the server is busy,} \\ 2, & \text{the server is during a normal service period at time } t \text{ and the server is free,} \\ 3, & \text{the server is during a normal service period at time } t \text{ and the server is busy.} \end{cases}$$

At time $t \geq 0$, we define the random variable $\xi(t)$ as follows: if $I(t) = 0$ or $I(t) = 2$ and $N(t) > 0$, $\xi(t)$ represents the elapsed retrial time; if $I(t) = 1$, $\xi(t)$ denotes the elapsed lower service time; if $I(t) = 3$, $\xi(t)$ stands for the elapsed normal service time. Therefore, the system can be described by Markov process $X(t) = \{I(t), N(t), \xi(t)\}$ with state space

$$\Omega = \{(0,0)\} \cup \{(1,0,x), x \geq 0\} \cup \{(2,0)\} \cup \{(3,0,x), x \geq 0\} \cup \{(i,n,x), i = 0,1,2,3, n \geq 1, x \geq 0\}.$$

Let $\{t_n; n = 1,2,\ldots\}$ be the sequence of epochs at which a normal service or a lower service completion occurs and $Y_n = \{I(t_n^+), N(t_n^+)\}$. Then the sequence of random variables $\{Y_n; n \geq 1\}$ forms an embedded Markov chain with state space $\{(0,0)\} \cup \{(i,k), i = 0,2, k \geq 1\}$.

3. Stable condition

To develop the transition probability matrix of $\{Y_n;\, n \geq 1\}$, we introduce a few definitions:
(1) Define

$$a_k = \int_0^\infty \frac{(\lambda x)^k}{k!} \mathrm{e}^{-\lambda x}\, \mathrm{d}G_b(x), \quad k \geq 0,$$

which represents the probability that k customers arrive during S_b, and the probability generating function of $\{a_k, k \geq 0\}$ is given by

$$A(z) \triangleq \sum_{k=0}^\infty a_k z^k = \int_0^\infty \mathrm{e}^{-\lambda(1-z)x}\, \mathrm{d}G_b(x) = \widetilde{G}_b\Big(\lambda(1-z)\Big), \quad 0 \leq z \leq 1.$$

We can also have

$$A(1) = 1, \quad A'(1) = \lambda \int_0^\infty x\, \mathrm{d}G_b(x) = \lambda\mu_1, \quad A''(1) = \lambda^2 \int_0^\infty x^2\, \mathrm{d}G_b(x) = \lambda^2\mu_2.$$

(2) Define

$$b_k = \int_0^\infty \frac{(\lambda x)^k}{k!} \mathrm{e}^{-\lambda x} \mathrm{e}^{-\theta x}\, \mathrm{d}G_v(x), \quad k \geq 0,$$

which explains the probability that $V \geq S_v$ and k customers arrive during S_v, and the probability generating function of $\{b_k, k \geq 0\}$ is given by

$$B(z) \triangleq \sum_{k=0}^\infty b_k z^k = \int_0^\infty \mathrm{e}^{-\left(\theta+\lambda(1-z)\right)x}\, \mathrm{d}G_v(x) = \widetilde{G}_v\Big(\theta + \lambda(1-z)\Big), \quad 0 \leq z \leq 1.$$

We can also get

$$B(1) = \widetilde{G}_v(\theta), \quad B'(1) = \lambda \int_0^\infty x \mathrm{e}^{-\theta x}\, \mathrm{d}G_v(x), \quad B''(1) = \lambda^2 \int_0^\infty x^2 \mathrm{e}^{-\theta x}\, \mathrm{d}G_v(x).$$

(3) Define

$$v_k = \int_0^\infty \frac{(\lambda x)^k}{k!} \mathrm{e}^{-\lambda x} \theta \mathrm{e}^{-\theta x} \overline{G}_v(x)\, \mathrm{d}x, \quad k \geq 0,$$

which represents the probability that $V < S_v$ and k customers arrive during V, and the probability generating function of $\{v_k, k \geq 0\}$ is given by

$$V(z) \triangleq \sum_{k=0}^{\infty} v_k z^k = \theta \overline{G}_v^* \Big(\theta + \lambda(1-z)\Big) = \frac{\theta}{\theta + \lambda(1-z)} \Big[1 - B(z)\Big], \quad 0 \leq z \leq 1.$$

We can also obtain

$$V(1) = 1 - \widetilde{G}_v(\theta), \quad V'(1) = \frac{\lambda}{\theta}\Big(1 - \widetilde{G}_v(\theta)\Big) - B'(1),$$
$$V''(1) = 2\Big(\frac{\lambda}{\theta}\Big)^2 \Big(1 - \widetilde{G}_v(\theta)\Big) - 2\frac{\lambda}{\theta} B'(1) - B''(1).$$

(4) Define

$$c_k = \sum_{j=0}^{k} v_j a_{k-j}, \quad k \geq 0,$$

which explains the probability that $V < S_v$ and k customers arrive during V plus S_b, and the probability generating function of $\{c_k, k \geq 0\}$ is given by

$$C(z) \triangleq \sum_{k=0}^{\infty} c_k z^k = V(z)A(z), \quad 0 \leq z \leq 1.$$

We can also derive

$$C(1) = V(1)A(1) = 1 - \widetilde{G}_v(\theta), \quad C'(1) = V'(1)A(1) + V(1)A'(1) = \lambda\Big(\mu_1 + \frac{1}{\theta}\Big)\Big(1 - \widetilde{G}_v(\theta)\Big) - B'(1),$$
$$\begin{aligned} C''(1) &= V''(1)A(1) + 2V'(1)A'(1) + V(1)A''(1) \\ &= \lambda^2\Big(\frac{2\mu_1}{\theta} + \frac{2}{\theta^2} + \mu_2\Big)\Big(1 - \widetilde{G}_v(\theta)\Big) - 2\lambda\Big(\mu_1 + \frac{1}{\theta}\Big)B'(1) - B''(1). \end{aligned}$$

Using the lexicographical sequence for the states, the transition probability matrix of $\{Y_n; n \geq 1\}$ can be written as the block-Jacobi matrix

$$P = \begin{pmatrix} W_0 & W_1 & W_2 & W_3 & \cdots \\ B_0 & A_1 & A_2 & A_3 & \cdots \\ & A_0 & A_1 & A_2 & \cdots \\ & & A_0 & A_1 & \cdots \\ & & & \ddots & \ddots \end{pmatrix},$$

where

$$W_0 = \frac{\lambda}{\lambda + \theta}(b_0 + c_0) + \frac{\theta}{\lambda + \theta} a_0, \quad W_k = \Big(\frac{\lambda}{\lambda + \theta}(b_k + q c_k), \frac{\lambda}{\lambda + \theta}\overline{q} c_k + \frac{\theta}{\lambda + \theta} a_k\Big), \quad k \geq 1,$$
$$B_0 = \begin{pmatrix} \widetilde{R}(\lambda)(b_0 + c_0) \\ \widetilde{R}(\lambda) a_0 \end{pmatrix}, \quad A_0 = \begin{pmatrix} \widetilde{R}(\lambda)(b_0 + q c_0) & \overline{q}\widetilde{R}(\lambda) c_0 \\ q\widetilde{R}(\lambda) a_0 & \overline{q}\widetilde{R}(\lambda) a_0 \end{pmatrix},$$
$$A_k = \begin{pmatrix} \widetilde{R}(\lambda)(b_k + q c_k) + [1 - \widetilde{R}(\lambda)](b_{k-1} + q c_{k-1}) & \overline{q}\widetilde{R}(\lambda) c_k + \overline{q}[1 - \widetilde{R}(\lambda)] c_{k-1} \\ q\widetilde{R}(\lambda) a_k + q[1 - \widetilde{R}(\lambda)] a_{k-1} & \overline{q}\widetilde{R}(\lambda) a_k + \overline{q}[1 - \widetilde{R}(\lambda)] a_{k-1} \end{pmatrix}, \quad k \geq 1.$$

We can easily check that

$$W_0 + \sum_{k=1}^{\infty} W_k e = 1, \quad B_0 + \sum_{k=1}^{\infty} A_k e = e, \quad \sum_{k=0}^{\infty} A_k e = e,$$

where $e = (1, 1)^T$.

Theorem 3.1. *The embedded Markov chain* $\{Y_n; n \geq 1\}$ *is ergodic if and only if* $\lambda\Big(\mu_1+\dfrac{q}{\theta}\Big)\Big(1-\widetilde{G}_v(\theta)\Big) < \widetilde{R}(\lambda)\Big[q+\overline{q}\Big(1-\widetilde{G}_v(\theta)\Big)\Big]$.

Proof. It is not difficult to see that $\{Y_n; n \geq 1\}$ is an irreducible and aperiodic Markov chain, so we just need to prove that $\{Y_n; n \geq 1\}$ is positive recurrent if and only if $\lambda\Big(\mu_1 + \dfrac{q}{\theta}\Big)\Big(1 - \widetilde{G}_v(\theta)\Big) < \widetilde{R}(\lambda)\Big[q+\overline{q}\Big(1-\widetilde{G}_v(\theta)\Big)\Big]$. We can have

$$A = \sum_{k=0}^{\infty} A_k = \begin{pmatrix} B(1)+qC(1) & \overline{q}C(1) \\ q & \overline{q} \end{pmatrix} = \begin{pmatrix} 1-\overline{q}C(1) & \overline{q}C(1) \\ q & \overline{q} \end{pmatrix},$$

and the invariant probability vector of matrix A is $\pi = (\pi_1, \pi_2)$, where

$$\pi_1 = \frac{q}{q+\overline{q}C(1)}, \pi_2 = \frac{\overline{q}C(1)}{q+\overline{q}C(1)}.$$

The vector β is defined by

$$\beta = \sum_{k=0}^{\infty} kA_k e.$$

And β is explicitly given by

$$\beta = \Big(1-\widetilde{R}(\lambda)+\lambda\Big(\mu_1+\frac{1}{\theta}\Big)\Big(1-\widetilde{G}_v(\theta)\Big),\ 1-\widetilde{R}(\lambda)+\lambda\mu_1\Big)^T.$$

It is clear from the Chapter 2 of Neuts [17], the embedded Markov chain $\{Y_n; n \geq 1\}$ is positive recurrent if and only if

$$\pi\beta < 1 \Longleftrightarrow \lambda\Big(\mu_1+\frac{q}{\theta}\Big)\Big(1-\widetilde{G}_v(\theta)\Big) < \widetilde{R}(\lambda)\Big[q+\overline{q}\Big(1-\widetilde{G}_v(\theta)\Big)\Big].$$

□

Since the arrival process is Poisson, using PASTA property, it can be showed from Burke's theorem (see [21], pp. 187–188) that the steady state probabilities of the Markov process $X(t)$ exist if and only if the stable condition $\lambda\Big(\mu_1+\dfrac{q}{\theta}\Big)\Big(1-\widetilde{G}_v(\theta)\Big) < \widetilde{R}(\lambda)\Big[q+\overline{q}\Big(1-\widetilde{G}_v(\theta)\Big)\Big]$ holds.

Now we define the limiting probabilities and limiting probability densities:

$$\begin{aligned}
P_{0,0} &= \lim_{t\to\infty} P(I(t)=0, N(t)=0),\\
P_{2,0} &= \lim_{t\to\infty} P(I(t)=2, N(t)=0),\\
P_{0,n}(x)\,\mathrm{d}x &= \lim_{t\to\infty} P(I(t)=0, N(t)=n, x\leq \xi(t) < x+\mathrm{d}x), \quad n\geq 1,\\
P_{1,n}(x)\,\mathrm{d}x &= \lim_{t\to\infty} P(I(t)=1, N(t)=n, x\leq \xi(t) < x+\mathrm{d}x), \quad n\geq 0,\\
P_{2,n}(x)\,\mathrm{d}x &= \lim_{t\to\infty} P(I(t)=2, N(t)=n, x\leq \xi(t) < x+\mathrm{d}x), \quad n\geq 1,\\
P_{3,n}(x)\,\mathrm{d}x &= \lim_{t\to\infty} P(I(t)=3, N(t)=n, x\leq \xi(t) < x+\mathrm{d}x), \quad n\geq 0.
\end{aligned}$$

4. Steady state analysis

By the method of supplementary variable technique, we obtain the following system of equations that govern the dynamics of the system

$$(\lambda+\theta)P_{0,0} = \int_0^{\infty} P_{1,0}(x)\eta(x)\,\mathrm{d}x + \int_0^{\infty} P_{3,0}(x)\mu(x)\,\mathrm{d}x, \tag{4.1}$$

$$\lambda P_{2,0} = \theta P_{0,0}, \tag{4.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}P_{0,n}(x) = -(\lambda+\theta+\alpha(x))P_{0,n}(x), \quad n \geq 1, \tag{4.3}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}P_{1,n}(x) = -(\lambda+\theta+\eta(x))P_{1,n}(x) + (1-\delta_{n,0})\lambda P_{1,n-1}(x), \quad n \geq 0, \tag{4.4}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}P_{2,n}(x) = -(\lambda+\alpha(x))P_{2,n}(x), \quad n \geq 1, \tag{4.5}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}P_{3,n}(x) = -(\lambda+\mu(x))P_{3,n}(x) + (1-\delta_{n,0})\lambda P_{3,n-1}(x), \quad n \geq 0, \tag{4.6}$$

where $\delta_{n,0}$ is the Kronecker's symbol. The boundary conditions are

$$P_{0,n}(0) = \int_0^\infty P_{1,n}(x)\eta(x)\,\mathrm{d}x + q\int_0^\infty P_{3,n}(x)\mu(x)\,\mathrm{d}x, \quad n \geq 1, \tag{4.7}$$

$$P_{1,n}(0) = \delta_{n,0}\lambda P_{0,0} + (1-\delta_{n,0})\lambda\int_0^\infty P_{0,n}(x)\,\mathrm{d}x + \int_0^\infty P_{0,n+1}(x)\alpha(x)\,\mathrm{d}x, \quad n \geq 0, \tag{4.8}$$

$$P_{2,n}(0) = \theta\int_0^\infty P_{0,n}(x)\,\mathrm{d}x + \overline{q}\int_0^\infty P_{3,n}(x)\mu(x)\,\mathrm{d}x, \quad n \geq 1, \tag{4.9}$$

$$P_{3,n}(0) = \delta_{n,0}\lambda P_{2,0} + \theta\int_0^\infty P_{1,n}(x)\,\mathrm{d}x + (1-\delta_{n,0})\lambda\int_0^\infty P_{2,n}(x)\,\mathrm{d}x + \int_0^\infty P_{2,n+1}(x)\alpha(x)\,\mathrm{d}x, \quad n \geq 0, \tag{4.10}$$

and the normalization condition is

$$P_{0,0} + P_{2,0} + \sum_{n=1}^\infty\left(\int_0^\infty P_{0,n}(x)\,\mathrm{d}x + \int_0^\infty P_{2,n}(x)\,\mathrm{d}x\right) + \sum_{n=0}^\infty\left(\int_0^\infty P_{1,n}(x)\,\mathrm{d}x + \int_0^\infty P_{3,n}(x)\,\mathrm{d}x\right) = 1. \tag{4.11}$$

By introducing the generating functions $P_i(x,z) = \sum_{n=b}^\infty P_{i,n}(x)z^n, i = 0,2, b = 1; i = 1,3, b = 0$, from equations (4.3)–(4.6), we can have

$$P_0(x,z) = P_0(0,z)\mathrm{e}^{-(\lambda+\theta)x}\overline{R}(x), \tag{4.12}$$

$$P_1(x,z) = P_1(0,z)\mathrm{e}^{-(\theta+\lambda(1-z))x}\overline{G}_v(x), \tag{4.13}$$

$$P_2(x,z) = P_2(0,z)\mathrm{e}^{-\lambda x}\overline{R}(x), \tag{4.14}$$

$$P_3(x,z) = P_3(0,z)\mathrm{e}^{-\lambda(1-z)x}\overline{G}_b(x). \tag{4.15}$$

From equations (4.1), (4.7)–(4.10), after some computations, we can obtain

$$(\lambda+\theta)P_{0,0} = P_{1,0}(0)b_0 + P_{3,0}(0)a_0, \tag{4.16}$$

$$P_0(0,z) = B(z)P_1(0,z) + qA(z)P_3(0,z) + \overline{q}P_{3,0}(0)a_0 - (\lambda+\theta)P_{0,0}, \tag{4.17}$$

$$zP_1(0,z) = \left(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)z\right)P_0(0,z) + \lambda P_{0,0}z, \tag{4.18}$$

$$P_2(0,z) = \theta\overline{R}^*(\lambda+\theta)P_0(0,z) + \overline{q}A(z)P_3(0,z) - \overline{q}P_{3,0}(0)a_0, \tag{4.19}$$

$$zP_3(0,z) = zV(z)P_1(0,z) + \left(\widetilde{R}(\lambda) + \lambda\overline{R}^*(\lambda)z\right)P_2(0,z) + \theta P_{0,0}z. \tag{4.20}$$

In the following, we consider three cases to obtain $P_0(0,z), P_1(0,z), P_2(0,z)$ and $P_3(0,z)$.

4.1. Case 1: $q = 0$

If $q = 0$, the model reduces to an M/G/1 retrial queue with general retrial times and single working vacation. The equations (4.17) and (4.19) become

$$P_0(0,z) = B(z)P_1(0,z) + P_{3,0}(0)a_0 - (\lambda+\theta)P_{0,0}, \tag{4.21}$$

$$P_2(0,z) = \theta\overline{R}^*(\lambda+\theta)P_0(0,z) + A(z)P_3(0,z) - P_{3,0}(0)a_0. \tag{4.22}$$

Using equations (4.21) and (4.18), we can get

$$P_0(0,z) = \frac{\lambda P_{0,0}B(z) + P_{3,0}(0)a_0 - (\lambda+\theta)P_{0,0}}{z - \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)z\Big)B(z)}z \triangleq \frac{\lambda P_{0,0}B(z) + P_{3,0}(0)a_0 - (\lambda+\theta)P_{0,0}}{f(z)}z, \tag{4.23}$$

$$P_1(0,z) = \Big(\frac{\widetilde{R}(\lambda+\theta)}{z} + \lambda\overline{R}^*(\lambda+\theta)\Big)P_0(0,z) + \lambda P_{0,0}. \tag{4.24}$$

From equations (4.22) and (4.20), we have

$$P_2(0,z) = \frac{\theta\overline{R}^*(\lambda+\theta)P_0(0,z) + C(z)P_1(0,z) - P_{3,0}(0)a_0 + \theta P_{0,0}A(z)}{z - \Big(\widetilde{R}(\lambda) + \lambda\overline{R}^*(\lambda)z\Big)A(z)}z \triangleq \frac{h_0(z)}{g_0(z)}z, \tag{4.25}$$

$$P_3(0,z) = V(z)P_1(0,z) + \Big(\frac{\widetilde{R}(\lambda)}{z} + \lambda\overline{R}^*(\lambda)\Big)P_2(0,z) + \theta P_{0,0}. \tag{4.26}$$

In order to obtain the relationship between $P_{0,0}$ and $P_{3,0}(0)$, we first give a lemma here.

Lemma 4.1. *The equation $f(z) = 0$ has the unique root $z = \alpha$ in the interval (0,1).*

Proof. Clearly,

$$f(0) = -\widetilde{R}(\lambda+\theta)B(0) = -\widetilde{R}(\lambda+\theta)\widetilde{G}_v(\lambda+\theta) < 0,$$
$$f(1) = 1 - \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)B(1) = 1 - \frac{\lambda+\theta\widetilde{R}(\lambda+\theta)}{\lambda+\theta}\widetilde{G}_v(\theta) > 0.$$

For any $0 < z < 1$, we have

$$f'(z) = 1 - \lambda\overline{R}^*(\lambda+\theta)B(z) - \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)z\Big)B'(z),$$
$$f''(z) = -2\lambda\overline{R}^*(\lambda+\theta)B'(z) - \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)z\Big)B''(z) < 0,$$

which means $f(z)$ is a concave function in the interval (0,1). Thus, $f(0) < 0$ and $f(1) > 0$ indicate that $f(z) = 0$ has the unique root $z = \alpha$ in the interval (0,1). □

From Lemma 4.1, the denominator of $P_0(0,z)$ is equal to 0 if $z = \alpha$, so does the numerator. Substituting $z = \alpha$ into the numerator of the right-hand side of (4.23), we have

$$P_{3,0}(0)a_0 = (\lambda+\theta)P_{0,0} - \lambda P_{0,0}B(\alpha).$$

Using the above equation, from equations (4.23)–(4.26), we can see that $P_0(0,z), P_1(0,z), P_2(0,z)$ and $P_3(0,z)$ will be expressed in terms of $P_{0,0}$, and $P_{0,0}$ is given in appendix.

Lemma 4.2. *If $q = 0$, some results which will be used in the generating functions of the number of customers are given as follows:*

$$f(1) = 1 - \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)B(1),$$
$$f'(1) = 1 - \lambda\overline{R}^*(\lambda+\theta)B(1) - \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)B'(1),$$
$$f''(1) = -2\lambda\overline{R}^*(\lambda+\theta)B'(1) - \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)B''(1),$$
$$P_0(0,1) = \frac{\lambda P_{0,0}B(1) - (\lambda+\theta)P_{0,0} + P_{3,0}(0)a_0}{f(1)},$$
$$P_0'(0,1) = \lim_{z\to1} P_0'(0,z) = \frac{\lambda P_{0,0}B'(1) - f'(1)P_0(0,1) + f(1)P_0(0,1)}{f(1)},$$
$$P_0''(0,1) = \lim_{z\to1} P_0''(0,z) = \frac{\lambda P_{0,0}B''(1) + 2\lambda P_{0,0}B'(1) - 2f'(1)P_0'(0,1) - f''(1)P_0(0,1)}{f(1)},$$
$$P_1(0,1) = \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)P_0(0,1) + \lambda P_{0,0},$$

$$P_1'(0,1)=\lim_{z\to 1}P_1'(0,z)=-\widetilde{R}(\lambda+\theta)P_0(0,1)+\Big(\widetilde{R}(\lambda+\theta)+\lambda\overline{R}^*(\lambda+\theta)\Big)P_0'(0,1),$$
$$P_1''(0,1)=\lim_{z\to 1}P_1''(0,z)=2\widetilde{R}(\lambda+\theta)P_0(0,1)-2\widetilde{R}(\lambda+\theta)P_0'(0,1)+\Big(\widetilde{R}(\lambda+\theta)+\lambda\overline{R}^*(\lambda+\theta)\Big)P_0''(0,1),$$
$$g_0(1)=0,\quad g_0'(1)=1-\lambda\overline{R}^*(\lambda)-A'(1)=\widetilde{R}(\lambda)-A'(1),\quad g_0''(1)=-2\lambda\overline{R}^*(\lambda)A'(1)-A''(1),$$
$$h_0(1)=\theta\overline{R}^*(\lambda+\theta)P_0(0,1)+C(1)P_1(0,1)-P_{3,0}(0)a_0+\theta P_{0,0}=0,$$
$$h_0'(1)=\theta\overline{R}^*(\lambda+\theta)P_0'(0,1)+C'(1)P_1(0,1)+C(1)P_1'(0,1)+\theta P_{0,0}A'(1),$$
$$h_0''(1)=\theta\overline{R}^*(\lambda+\theta)P_0''(0,1)+C''(1)P_1(0,1)+2C'(1)P_1'(0,1)+C(1)P_1''(0,1)+\theta P_{0,0}A''(1),$$
$$P_2(0,1)=\frac{h_0'(1)}{g_0'(1)},\quad P_2'(0,1)=\lim_{z\to 1}P_2'(0,z)=\frac{h_0''(1)g_0'(1)-h_0'(1)g_0''(1)+2h_0'(1)g_0'(1)}{2(g_0'(1))^2},$$
$$P_3(0,1)=V(1)P_1(0,1)+P_2(0,1)+\theta P_{0,0},$$
$$P_3'(0,1)=\lim_{z\to 1}P_3'(0,z)=V'(1)P_1(0,1)+V(1)P_1'(0,1)-\widetilde{R}(\lambda)P_2(0,1)+P_2'(0,1).$$

Proof. This lemma can be obtained by some tedious algebraic manipulations, and the proof is too long but straightforward. For space consideration, we omit it here. □

Remark 4.3. Since $g_0(1)$=0 and $h_0(1)$=0, we need to apply the first L'Hospital rule to compute $P_2(0,1)$, and use the L'Hospital rule twice to calculate $P_2'(0,1)$.

4.2. Case 2: $q = 1$

If $q=1$, the equations (4.17) and (4.19) become

$$P_0(0,z)=B(z)P_1(0,z)+A(z)P_3(0,z)-(\lambda+\theta)P_{0,0},\tag{4.27}$$
$$P_2(0,z)=\theta\overline{R}^*(\lambda+\theta)P_0(0,z).\tag{4.28}$$

Taking (4.28) into (4.20), and then inserting (4.20) and (4.18) into (4.27), we can get

$$\begin{aligned}P_0(0,z)&=\frac{\lambda\Big(B(z)+C(z)\Big)P_{0,0}+\theta A(z)P_{0,0}-(\lambda+\theta)P_{0,0}}{f(z)-\Big(\widetilde{R}(\lambda+\theta)+\lambda\overline{R}^*(\lambda+\theta)z\Big)C(z)-\Big(\widetilde{R}(\lambda)+\lambda\overline{R}^*(\lambda)z\Big)\theta\overline{R}^*(\lambda+\theta)A(z)}z\\&\triangleq\frac{M_1(z)}{N_1(z)}z.\end{aligned}\tag{4.29}$$

Clearly, once $P_0(0,z)$ is derived, $P_1(0,z),P_2(0,z)$ and $P_3(0,z)$ can be obtained from equations (4.18), (4.28) and (4.20). We can also see that $P_0(0,z),P_1(0,z),P_2(0,z)$ and $P_3(0,z)$ will be expressed in terms of $P_{0,0}$, and $P_{0,0}$ is given in appendix.

Lemma 4.4. *If $q=1$, some results which will be used in the generating functions of the number of customers are given as follows:*

$$M_1(1)=0,\quad M_1'(1)=\lambda\Big(B'(1)+C'(1)\Big)P_{0,0}+\theta A'(1)P_{0,0},$$
$$M_1''(1)=\lambda\Big(B''(1)+C''(1)\Big)P_{0,0}+\theta A''(1)P_{0,0},$$
$$N_1(1)=f(1)-\Big(\widetilde{R}(\lambda+\theta)+\lambda\overline{R}^*(\lambda+\theta)\Big)C(1)-\theta\overline{R}^*(\lambda+\theta)=0,$$
$$\begin{aligned}N_1'(1)=&f'(1)-\lambda\overline{R}^*(\lambda+\theta)C(1)-\Big(\widetilde{R}(\lambda+\theta)+\lambda\overline{R}^*(\lambda+\theta)\Big)C'(1)\\&-\lambda\overline{R}^*(\lambda)\theta\overline{R}^*(\lambda+\theta)-\theta\overline{R}^*(\lambda+\theta)A'(1),\end{aligned}$$
$$\begin{aligned}N_1''(1)=&f''(1)-2\lambda\overline{R}^*(\lambda+\theta)C'(1)-\Big(\widetilde{R}(\lambda+\theta)+\lambda\overline{R}^*(\lambda+\theta)\Big)C''(1)\\&-2\lambda\overline{R}^*(\lambda)\theta\overline{R}^*(\lambda+\theta)A'(1)-\theta\overline{R}^*(\lambda+\theta)A''(1),\end{aligned}$$
$$P_0(0,1)=\frac{M_1'(1)}{N_1'(1)},\quad P_0'(0,1)=\lim_{z\to 1}P_0'(0,z)=\frac{M_1''(1)N_1'(1)-M_1'(1)N_1''(1)+2M_1'(1)N_1'(1)}{2(N_1'(1))^2},$$

$$P_1(0,1)=\Big(\widetilde{R}(\lambda+\theta)+\lambda\overline{R}^*(\lambda+\theta)\Big)P_0(0,1)+\lambda P_{0,0},$$
$$P_1'(0,1)=\lim_{z\to1}P_1'(0,z)=-\widetilde{R}(\lambda+\theta)P_0(0,1)+\Big(\widetilde{R}(\lambda+\theta)+\lambda\overline{R}^*(\lambda+\theta)\Big)P_0'(0,1),$$
$$P_2(0,1)=\theta\overline{R}^*(\lambda+\theta)P_0(0,1),\quad P_2'(0,1)=\lim_{z\to1}P_2'(0,z)=\theta\overline{R}^*(\lambda+\theta)P_0'(0,1),$$
$$P_3(0,1)=V(1)P_1(0,1)+P_2(0,1)+\theta P_{0,0},$$
$$P_3'(0,1)=\lim_{z\to1}P_3'(0,z)=V'(1)P_1(0,1)+V(1)P_1'(0,1)-\widetilde{R}(\lambda)P_2(0,1)+P_2'(0,1).$$

Proof. This lemma can be obtained by some tedious algebraic manipulations, and the proof is too long but straightforward. For space consideration, we omit it here. □

Remark 4.5. Since $M_1(1)$=0 and $N_1(1)$=0, we need to apply the first L'Hospital rule to compute $P_0(0,1)$, and use the L'Hospital rule twice to calculate $P_0'(0,1)$.

4.3. Case 3: $0<q<1$

Using equations (4.19) and (4.20), we can get

$$P_2(0,z)=\frac{\theta\overline{R}^*(\lambda+\theta)P_0(0,z)+\overline{q}C(z)P_1(0,z)-\overline{q}P_{3,0}(0)a_0+\overline{q}\theta A(z)P_{0,0}}{z-\overline{q}\Big(\widetilde{R}(\lambda)+\lambda\overline{R}^*(\lambda)z\Big)A(z)}z\triangleq\frac{h(z)}{g(z)}z. \tag{4.30}$$

Taking (4.30) into (4.20), and then inserting (4.20) and (4.18) into (4.17), after some manipulations, we can get

$$P_0(0,z)=\frac{\Big(\lambda B(z)-\lambda-\theta\Big)g(z)P_{0,0}+\Big(\lambda C(z)+\theta A(z)\Big)qzP_{0,0}+\Big(g(z)-qz\Big)P_{3,0}(0)a_0}{g(z)f(z)-q\Big(\widetilde{R}(\lambda+\theta)+\lambda\overline{R}^*(\lambda+\theta)z\Big)C(z)z-q\Big(\widetilde{R}(\lambda)+\lambda\overline{R}^*(\lambda)z\Big)\theta\overline{R}^*(\lambda+\theta)A(z)z}z$$
$$\triangleq\frac{M(z)}{N(z)}z. \tag{4.31}$$

Clearly, once $P_0(0,z)$ is derived, $P_1(0,z),P_2(0,z)$ and $P_3(0,z)$ can be obtained from equations (4.18), (4.30) and (4.20). Next, we will give the relationship between $P_{0,0}$ and $P_{3,0}(0)$, and a lemma is first given here.

Lemma 4.6. *If $0<q<1$, the equation $g(z)=0$ has the unique root $z=\beta$ in the interval (0,1).*

Proof. Clearly,

$$g(0)=-\overline{q}\widetilde{R}(\lambda)A(0)=-\overline{q}\widetilde{R}(\lambda)\widetilde{G}_b(\lambda)<0,$$
$$g(1)=1-\overline{q}=q>0.$$

For any $0<z<1$, we have

$$g'(z)=1-\overline{q}\lambda\overline{R}^*(\lambda)A(z)-\overline{q}\Big(\widetilde{R}(\lambda)+\lambda\overline{R}^*(\lambda)z\Big)A'(z),$$
$$g''(z)=-2\overline{q}\lambda\overline{R}^*(\lambda)A'(z)-\overline{q}\Big(\widetilde{R}(\lambda)+\lambda\overline{R}^*(\lambda)z\Big)A''(z)<0,$$

which means $g(z)$ is a concave function in the interval (0,1). Thus, $g(0)<0$ and $g(1)>0$ indicate that $g(z)=0$ has the unique root $z=\beta$ in the interval (0,1). □

Furthermore, we can easily get $N(0)>0$ and $N(\beta)<0$, so $N(z)=0$ has a root $z=\gamma$ in the interval (0,β), which means that the numerator of $P_0(0,z)$ is equal to zero when $z=\gamma$, *i.e.*, $M(\gamma)=0$, and we can have

$$P_{3,0}(0)a_0=\frac{\Big(\lambda B(\gamma)-\lambda-\theta\Big)g(\gamma)+\Big(\lambda C(\gamma)+\theta A(\gamma)\Big)q\gamma}{q\gamma-g(\gamma)}P_{0,0}\triangleq L(\gamma)P_{0,0}.$$

Using the above equation, we can see that $P_0(0,z),P_1(0,z),P_2(0,z)$ and $P_3(0,z)$ will be expressed in terms of $P_{0,0}$, and $P_{0,0}$ is given in appendix.

Lemma 4.7. *If $0 < q < 1$, some results which will be used in the generating functions of the number of customers are given as follows:*

$$g(1) = q, \quad g'(1) = 1 - \overline{q}\lambda\overline{R}^*(\lambda) - \overline{q}A'(1), \quad g''(1) = -2\overline{q}\lambda\overline{R}^*(\lambda)A'(1) - \overline{q}A''(1), \quad M(1) = 0,$$

$$M'(1) = \lambda B'(1)qP_{0,0} + \Big(\lambda B(1) - \lambda - \theta\Big)g'(1)P_{0,0} + \Big(\lambda C'(1) + \theta A'(1)\Big)qP_{0,0} + \Big(\lambda C(1) + \theta\Big)qP_{0,0} + \Big(g'(1) - q\Big)P_{3,0}(0)a_0,$$

$$M''(1) = \lambda B''(1)qP_{0,0} + 2\lambda B'(1)g'(1)P_{0,0} + \Big(\lambda B(1) - \lambda - \theta\Big)g''(1)P_{0,0} + \Big(\lambda C''(1) + \theta A''(1)\Big)qP_{0,0} + 2\Big(\lambda C'(1) + \theta A'(1)\Big)qP_{0,0} + g''(1)P_{3,0}(0)a_0,$$

$$N(1) = qf(1) - q\Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)C(1) - q\theta\overline{R}^*(\lambda+\theta) = 0,$$

$$N'(1) = g'(1)f(1) + qf'(1) - q\lambda\overline{R}^*(\lambda+\theta)C(1) - q\Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)\Big(C'(1) + C(1)\Big) - q\lambda\overline{R}^*(\lambda)\theta\overline{R}^*(\lambda+\theta) - q\theta\overline{R}^*(\lambda+\theta)\Big(A'(1) + 1\Big),$$

$$N''(1) = g''(1)f(1) + 2g'(1)f'(1) + qf''(1) - 2q\lambda\overline{R}^*(\lambda+\theta)\Big(C'(1) + C(1)\Big) - q\Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big) \times \Big(C''(1) + 2C'(1)\Big) - 2q\lambda\overline{R}^*(\lambda)\theta\overline{R}^*(\lambda+\theta)\Big(A'(1) + 1\Big) - q\theta\overline{R}^*(\lambda+\theta)\Big(A''(1) + 2A'(1)\Big),$$

$$P_0(0,1) = \frac{M'(1)}{N'(1)}, \quad P_0'(0,1) = \lim_{z\to1} P_0'(0,z) = \frac{M''(1)N'(1) - M'(1)N''(1) + 2M'(1)N'(1)}{2(N'(1))^2},$$

$$P_1(0,1) = \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)P_0(0,1) + \lambda P_{0,0},$$

$$P_1'(0,1) = \lim_{z\to1} P_1'(0,z) = -\widetilde{R}(\lambda+\theta)P_0(0,1) + \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)P_0'(0,1),$$

$$h(1) = \theta\overline{R}^*(\lambda+\theta)P_0(0,1) + \overline{q}C(1)P_1(0,1) - \overline{q}P_{3,0}(0)a_0 + \overline{q}\theta P_{0,0},$$

$$h'(1) = \theta\overline{R}^*(\lambda+\theta)P_0'(0,1) + \overline{q}C'(1)P_1(0,1) + \overline{q}C(1)P_1'(0,1) + \overline{q}\theta A'(1)P_{0,0},$$

$$P_2(0,1) = \frac{h(1)}{q}, \quad P_2'(0,1) = \lim_{z\to1} P_2'(0,z) = \frac{\Big(h'(1) + h(1)\Big)q - h(1)g'(1)}{q^2},$$

$$P_3(0,1) = V(1)P_1(0,1) + P_2(0,1) + \theta P_{0,0},$$

$$P_3'(0,1) = \lim_{z\to1} P_3'(0,z) = V'(1)P_1(0,1) + V(1)P_1'(0,1) - \widetilde{R}(\lambda)P_2(0,1) + P_2'(0,1).$$

Proof. This lemma can be obtained by some tedious algebraic manipulations, and the proof is too long but straightforward. For space consideration, we omit it here. ☐

Remark 4.8. Since $M(1)$=0 and $N(1)$=0, we need to apply the first L'Hospital rule to compute $P_0(0,1)$, and use the L'Hospital rule twice to calculate $P_0'(0,1)$.

5. Performance measure

In the above section, we have obtained $P_i(0,z), P_i(0,1)$ and $P_i'(0,1), i = 0,1,2,3$. Using these expressions, we can get some interesting performance measures. Define the marginal generating functions $\Phi_i(z) = \int_0^\infty P_i(x,z)dx, i = 0,1,2,3$, we can have the following theorem.

Theorem 5.1. *If $\lambda\Big(\mu_1 + \frac{q}{\theta}\Big)\Big(1 - \widetilde{G}_v(\theta)\Big) < \widetilde{R}(\lambda)\Big[q + \overline{q}\Big(1 - \widetilde{G}_v(\theta)\Big)\Big]$, the generating functions of the number of customers in the orbit in different states can be written as follows:*

$$\Phi_0(z) = P_0(0,z)\overline{R}^*(\lambda+\theta),$$

$$\Phi_1(z) = P_1(0,z)\frac{V(z)}{\theta},$$

$$\Phi_2(z) = P_2(0,z)\overline{R}^*(\lambda),$$

$$\Phi_3(z) = P_3(0,z)\frac{1-A(z)}{\lambda(1-z)},$$

where $P_{0,0}$ *is determined by the normalization condition*

$$P_{0,0} + P_{2,0} + \Phi_0(1) + \Phi_1(1) + \Phi_2(1) + \Phi_3(1) = 1.$$

Proof. Using equations (4.12)–(4.15), we can obtain

$$\begin{aligned}
\Phi_0(z) &= \int_0^\infty P_0(0,z)\mathrm{e}^{-(\lambda+\theta)x}\overline{R}(x)\,\mathrm{d}x = P_0(0,z)\overline{R}^*(\lambda+\theta),\\
\Phi_1(z) &= \int_0^\infty P_1(0,z)\mathrm{e}^{-(\theta+\lambda(1-z))x}\overline{G}_v(x)\,\mathrm{d}x = P_1(0,z)\overline{G}_v^*\Big(\theta+\lambda(1-z)\Big) = P_1(0,z)\frac{V(z)}{\theta},\\
\Phi_2(z) &= \int_0^\infty P_2(0,z)\mathrm{e}^{-\lambda x}\overline{R}(x)\,\mathrm{d}x = P_2(0,z)\overline{R}^*(\lambda),\\
\Phi_3(z) &= \int_0^\infty P_3(0,z)\mathrm{e}^{-\lambda(1-z)x}\overline{G}_b(x)\,\mathrm{d}x = P_3(0,z)\overline{G}_b^*\Big(\lambda(1-z)\Big) = P_3(0,z)\frac{1-A(z)}{\lambda(1-z)}.
\end{aligned}$$

Clearly, the normalization condition is

$$P_{0,0} + P_{2,0} + \Phi_0(1) + \Phi_1(1) + \Phi_2(1) + \Phi_3(1) = 1.$$

□

The probability generating function of the number of customers in the orbit is given by

$$\Phi(z) = P_{0,0} + P_{2,0} + \Phi_0(z) + \Phi_1(z) + \Phi_2(z) + \Phi_3(z).$$

The probability generating function of the number of customers in the system is given by

$$\widetilde{\Phi}(z) = P_{0,0} + P_{2,0} + \Phi_0(z) + z\Phi_1(z) + \Phi_2(z) + z\Phi_3(z).$$

The probability that the server is busy is

$$P_b = \Phi_1(1) + \Phi_3(1) = P_1(0,1)\frac{V(1)}{\theta} + P_3(0,1)\frac{A'(1)}{\lambda}\cdot$$

The probability that the server is free is

$$P_f = P_{0,0} + P_{2,0} + \Phi_0(1) + \Phi_2(1) = P_{0,0} + P_{2,0} + P_0(0,1)\overline{R}^*(\lambda+\theta) + P_2(0,1)\overline{R}^*(\lambda) = 1 - P_b.$$

The probability that the server is in a working vacation period is given by

$$P_w = P_{0,0} + \Phi_0(1) + \Phi_1(1) = P_{0,0} + P_0(0,1)\overline{R}^*(\lambda+\theta) + P_1(0,1)\frac{V(1)}{\theta}\cdot$$

The probability that the server is in a normal period is given by

$$P_n = P_{2,0} + \Phi_2(1) + \Phi_3(1) = P_{2,0} + P_2(0,1)\overline{R}^*(\lambda) + P_3(0,1)\frac{A'(1)}{\lambda} = 1 - P_w.$$

Let $E[L_j]$ denote the average number of customers in the orbit when the server's state is $i, i = 0,1,2,3$.
From Theorem 5.1, after some calculations we can get

$$\begin{aligned}
E[L_0] &= \lim_{z\to 1}\Phi_0'(z) = P_0'(0,1)\overline{R}^*(\lambda+\theta),\\
E[L_1] &= \lim_{z\to 1}\Phi_1'(z) = P_1'(0,1)\frac{V(1)}{\theta} + P_1(0,1)\frac{V'(1)}{\theta},\\
E[L_2] &= \lim_{z\to 1}\Phi_2'(z) = P_2'(0,1)\overline{R}^*(\lambda),\\
E[L_3] &= \lim_{z\to 1}\Phi_3'(z) = P_3'(0,1)\frac{A'(1)}{\lambda} + P_3(0,1)\frac{A''(1)}{2\lambda}\cdot
\end{aligned}$$

Therefore, the mean orbit length ($E[L]$) is given by

$$E[L]=\lim_{z\to 1}\Phi'(z)=E[L_0]+E[L_1]+E[L_2]+E[L_3].$$

And the mean system length ($E[\widetilde{L}]$) is derived as

$$E[\widetilde{L}]=\lim_{z\to 1}\widetilde{\Phi}'(z)=E[L]+\Phi_1(1)+\Phi_3(1)=E[L]+P_b.$$

Let $E[W]$ ($E[\widetilde{W}]$) be the expected waiting (sojourn) time of a customer in the orbit (system), using Little's formula, we can obtain

$$E[W]=\frac{E[L]}{\lambda},\quad E[\widetilde{W}]=\frac{E[\widetilde{L}]}{\lambda}\cdot$$

Theorem 5.2. *If* $\lambda\left(\mu_1+\frac{q}{\theta}\right)\left(1-\widetilde{G}_v(\theta)\right)<\widetilde{R}(\lambda)\left[q+\overline{q}\left(1-\widetilde{G}_v(\theta)\right)\right]$, *the probability generating function of the steady state distribution of the number of customers in the system at a departure epoch is given by*

$$\Pi(z)=\frac{P_1(0,z)B(z)+P_3(0,z)A(z)}{P_1(0,1)B(1)+P_3(0,1)}. \tag{5.1}$$

Proof. Following the argument of PASTA [22], we state that a departing customer will see n customers in the orbit just after a departure if and only if there are $n+1$ customers in the system (or n customers in the orbit) just before the departure. We denote $\{\pi_n;n\geq 0\}$ as the probability that there are n customers in the orbit at a departure epoch, then for $n\geq 0$, we can get

$$\pi_n=K\int_0^\infty P_{1,n}(x)\eta(x)\,\mathrm{d}x+K\int_0^\infty P_{3,n}(x)\mu(x)\,\mathrm{d}x,$$

where K is the normalizing constant. Define probability generating function $\Pi(z)=\sum_{n=0}^\infty \pi_n z^n$, we can obtain

$$\Pi(z)=K\int_0^\infty P_1(x,z)\eta(x)\,\mathrm{d}x+K\int_0^\infty P_3(x,z)\mu(x)\,\mathrm{d}x=KP_1(0,z)B(z)+KP_3(0,z)A(z). \tag{5.2}$$

Using the normalization condition $\Pi(1)$=1, we have

$$K=\frac{1}{P_1(0,1)B(1)+P_3(0,1)}. \tag{5.3}$$

Hence (5.1) follows by inserting (5.3) into (5.2). □

6. Numerical results

In this section, we present some numerical examples to illustrate the effect of the varying parameters on the mean orbit length $E[L]$. We consider two different retrial time distributions, exponential distribution with LST $\widetilde{R}(s)=\frac{\alpha}{s+\alpha}$ and Erlang distribution of order 2 with LST $\widetilde{R}(s)=(\frac{\alpha}{s+\alpha})^2$. Moreover, a cost minimization problem is also considered. For the simplicity purpose, it is assumed that the normal service time (the lower service time) is governed by an exponential distribution with parameter μ (η). Under the stable condition, all the computations are done by developing program in Matlab software and the values of parameters are arbitrarily chosen as $\lambda=1.5$, $\mu=1/\mu_1=5$, $\eta=1/\eta_1=1$, $\alpha=8$, $\theta=2$ and $q=0.5$, unless they are considered as variables in the respective figures.

6.1. Sensitivity analysis

From Figures 1 and 2, it is obvious that $E[L]$ decreases evidently as the values of η increase. When $\eta<\mu$, as expected, increasing θ decreases the value of $E[L]$. The effect of η on $E[L]$ is more obvious when θ is smaller, this is due to the fact that the expected vacation time is $1/\theta$. An especial case is $\eta\to\mu$, *i.e.*, the lower service rate equals to the normal service rate, it can be observed that θ has no

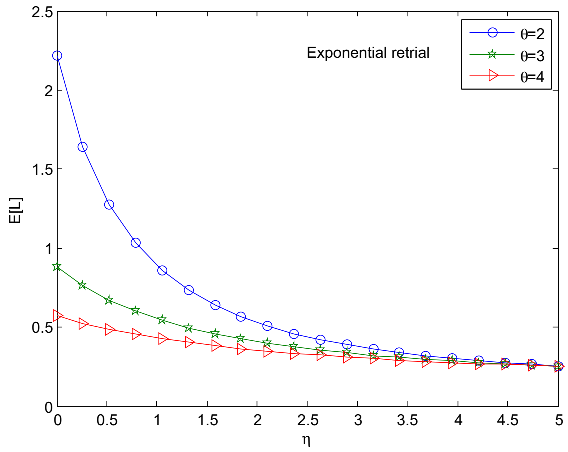

FIGURE 1. The effect of η on $E[L]$ for different values of θ.

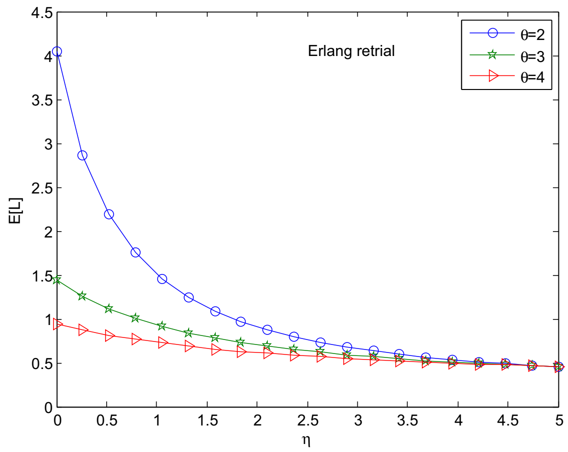

FIGURE 2. The effect of η on $E[L]$ for different values of θ.

effect on $E[L]$. Another extreme case is η=0, *i.e.*, the server cannot provide service in a vacation period, and θ has a noticeable effect on $E[L]$ which cannot be ignored.

In our model, if the orbit is not empty at a service completion instant in the normal period, the server can take a working vacation with probability q. Thus, as shown in Figures 3 and 4, when $\eta < \mu$, $E[L]$ increases with increasing values of q, but the effect of q on $E[L]$ is not obvious for a larger value of η. An especial case is $\eta = \mu$, and our model reduces to an M/G/1 queue without vacation, we can find that probability q has no effect on $E[L]$.

Figures 5 and 6 illustrate that $E[L]$ decreases as α increases, this is because that with the increment of α, the mean retrial time decreases. And the smaller the mean retrial time is, the bigger the probability that the server is busy is, which decreases the value of $E[L]$. Since $\eta < \mu$, the effect of α on $E[L]$ is more obvious when q is larger. Moreover, it can also be observed that $E[L]$ decreases with decreasing values of q.

Furthermore, with the same value of parameter α, the mean retrial time with exponential distribution is shorter than that with Erlang distribution. Thus, under the same condition, from Figures 1 to 6, we can see that the value of $E[L]$ with exponential retrial time is smaller than that with Erlang retrial time.

6.2. Cost analysis

In practice, queueing managers are always interested in minimizing operating cost of unit time. In this subsection, we establish a cost function to search for the optimal probability q, so as to minimize the expected operating cost function per unit time.

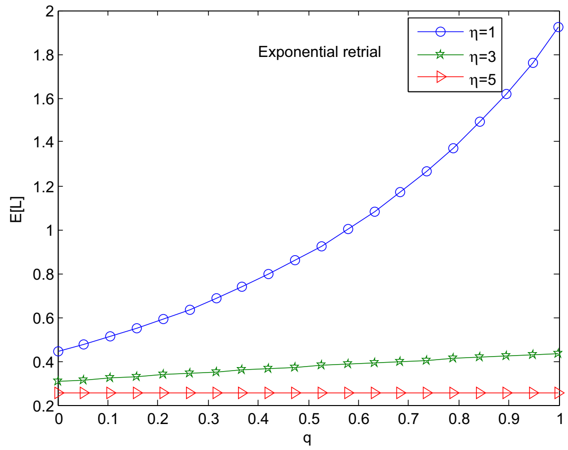

FIGURE 3. The effect of q on $E[L]$ for different values of η.

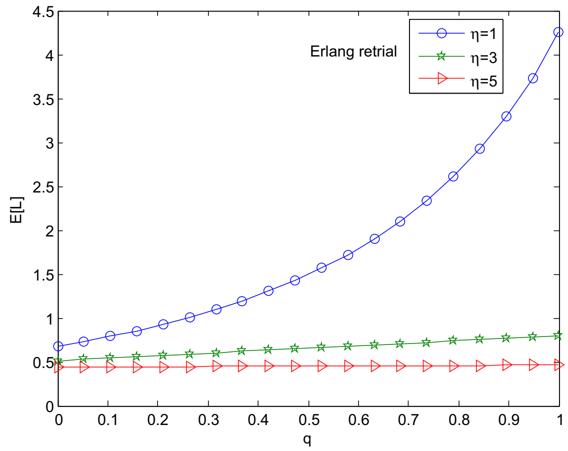

FIGURE 4. The effect of q on $E[L]$ for different values of η.

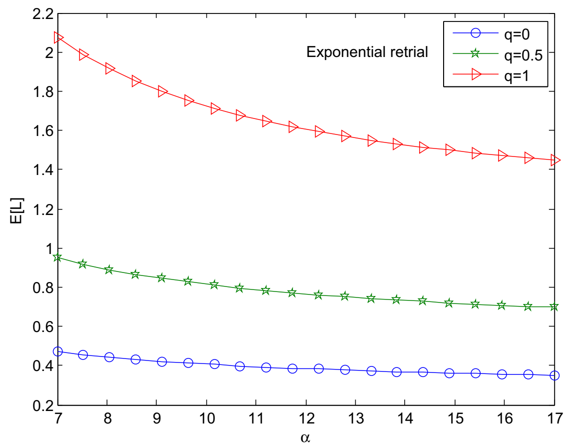

FIGURE 5. The effect of α on $E[L]$ for different values of q.

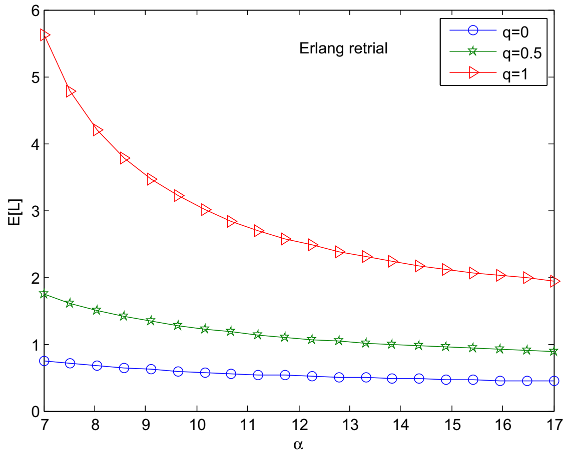

FIGURE 6. The effect of α on $E[L]$ for different values of q.

Define the following cost elements:
C_L=cost per unit time for each customer present in the orbit;
C_n=fixed cost per unit time when the server is in a normal period;
C_w=fixed cost per unit time when the server is in a working vacation period;
C_μ=cost per customer served by the normal service rate μ;
C_η=cost per customer served by the lower service rate η.
Based on the definitions of each cost element listed above, the expected operating cost function per unit time is given by

$$\min_q : f(q) = C_L E[L] + C_n P_n + C_w P_w + C_\mu \mu + C_\eta \eta.$$

Because the expected operating cost function per unit time is highly non-linear and complex, it is not easy to get the derivative of it. Hence in the following two examples are provided with assumption that $C_L = 3$, $C_n = 50$, $C_w = 23$, $C_\mu = 150$, $C_\eta = 80$, and we use the parabolic method to find the optimum value of q, say q^*. The essence of the parabolic method is to generate a quadratic function through the evaluated points in each iteration, and the objective function $f(x)$ is approximated by the quadratic function in generating an estimate of the optimum value. According to the polynomial approximation theory, the unique optimum of the quadratic function agreeing with $f(x)$ at 3-point pattern $\{x_0, x_1, x_2\}$ occurs at

$$\bar{x} = \frac{1}{2}\frac{f(x_0)(x_1^2 - x_2^2) + f(x_1)(x_2^2 - x_0^2) + f(x_2)(x_0^2 - x_1^2)}{f(x_0)(x_1 - x_2) + f(x_1)(x_2 - x_0) + f(x_2)(x_0 - x_1)}. \tag{6.1}$$

The steps of the parabolic method are given as follows:

Step 1: Choose a starting 3-point pattern $\{x_0, x_1, x_2\}$ along with a stopping tolerance ε, and initialize the iteration counter $i = 0$.

Step 2: Compute $\bar{x}$ according to (6.1). If $|\bar{x} - x_1| \le \varepsilon$, stop and report approximate optimum solution $\bar{x}$.

Step 3: If $\bar{x} \le x_1$, go to Step 4. If $\bar{x} > x_1$, go to Step 5.

Step 4: If $f(x_1)$ is less than $f(\bar{x})$, update $\bar{x} \to x_0$. Otherwise, replace $x_1 \to x_2, \bar{x} \to x_1$. Either way, advance $i = i + 1$, and return to Step 2.

Step 5: If $f(x_1)$ is less than $f(\bar{x})$, update $\bar{x} \to x_2$. Otherwise, replace $x_1 \to x_0, \bar{x} \to x_1$. Either way, advance $i = i + 1$, and return to Step 2.

Figures 7 and 8 plot the curve of the cost function with the change of q for an M/M/1 retrial queue, where the retrial time follows an exponential distribution with parameter $\alpha = 4$ and an Erlang distribution of order 2 with parameter $\alpha = 8$, respectively. Clearly, there is an optimal probability q to make the cost minimize. Using the parabolic method and the error is controlled by $\varepsilon = 10^{-4}$. After five iterations, Table 1 shows that the minimum expected operating cost per unit time converges to the solution $q^* = 0.55446$ with a value $f(q^*) = 872.26091$. From Table 2, after three iterations, we have the solution $q^* = 0.46970$ with $f(q^*) = 872.83711$.

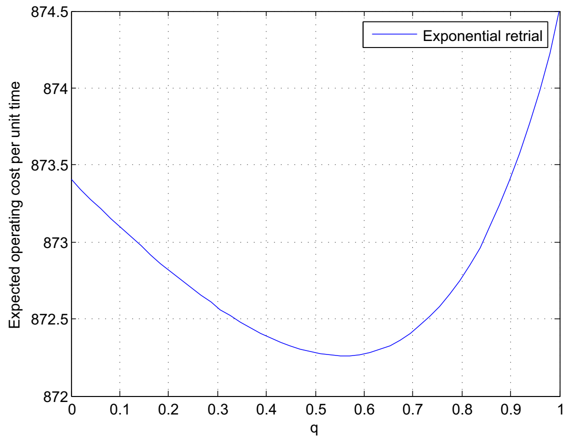

FIGURE 7. The effect of q on the expected operating cost per unit time.

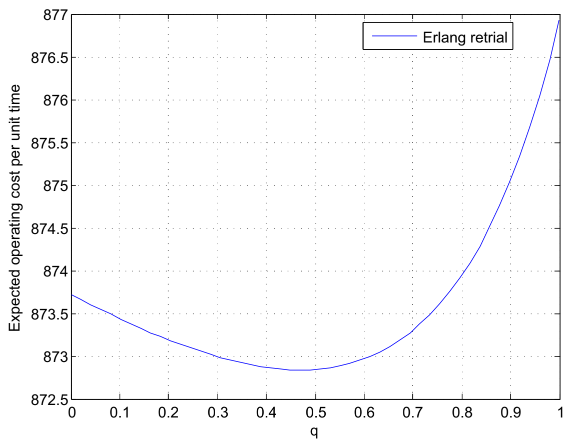

FIGURE 8. The effect of q on the expected operating cost per unit time.

TABLE 1. The parabolic method in searching the optimum solution of an M/M/1 queue (exponential retrial).

Iterations	x_0	x_1	x_2	$f(x_0)$	$f(x_1)$	$f(x_2)$	$\bar{x}$	$f(\bar{x})$	Tolerance
0	0.40000	0.50000	0.70000	872.38888	872.27864	872.41774	0.54197	872.26192	0.04197
1	0.50000	0.54197	0.70000	872.27864	872.26192	872.41774	0.54976	872.26106	0.00779
2	0.54197	0.54976	0.70000	872.26192	872.26106	872.41774	0.55336	872.26092	0.00359
3	0.54976	0.55336	0.70000	872.26106	872.26092	872.41774	0.55424	872.26091	0.00088
4	0.55336	0.55424	0.70000	872.26092	872.26091	872.41774	0.55452	872.26091	0.00028
5	0.55424	0.55452	0.70000	872.26091	872.26091	872.41774	0.55446	872.26091	0.00006

7. CONCLUSION

This paper investigates an M/G/1 retrial queue with general retrial times, and working vacation is controlled by Bernoulli schedule. The server may begin a working vacation at a service completion instant even if the system is not empty. Using the embedded Markov chain, we obtain the condition of stability. By applying the supplementary variable technique, we discuss the generating functions for different values of probability q. Various important performance measures are also obtained. Finally, we present some numerical examples and consider a cost minimization problem. For future research, using the same method, one can deal with a similar model but with batch arrival.

TABLE 2. The parabolic method in searching the optimum solution of an M/M/1 queue (Erlang retrial).

Iterations	x_0	x_1	x_2	$f(x_0)$	$f(x_1)$	$f(x_2)$	$\bar{x}$	$f(\bar{x})$	Tolerance
0	0.40000	0.50000	0.60000	872.86627	872.84328	872.96668	0.46571	872.83722	0.03429
1	0.40000	0.46571	0.50000	872.86627	872.83722	872.84328	0.46857	872.83712	0.00287
2	0.46571	0.46857	0.50000	872.83722	872.83712	872.84328	0.46963	872.83711	0.00106
3	0.46857	0.46963	0.50000	872.83712	872.83711	872.84328	0.46970	872.83711	0.00007

APPENDIX A

Here we give the expression of $P_{0,0}$ for three different cases, and the proof is omitted.

Case 1: $q = 0$

Define

$$\begin{aligned}
K_0(0) &= \frac{\lambda B(1) - \lambda B(\alpha)}{f(1)}, \\
\kappa_0 &= \frac{\lambda B'(1) - f'(1)K_0(0) + f(1)K_0(0)}{f(1)}, \\
K_0(1) &= \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)K_0(0) + \lambda, \\
\kappa_1 &= -\widetilde{R}(\lambda+\theta)K_0(0) + \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)\kappa_0, \\
K_0(2) &= \frac{\theta\overline{R}^*(\lambda+\theta)\kappa_0 + C'(1)K_0(1) + C(1)\kappa_1 + \theta A'(1)}{g_0'(1)}, \\
K_0(3) &= V(1)K_0(1) + K_0(2) + \theta.
\end{aligned}$$

$P_{0,0}$ is given by

$$P_{0,0} = \Big[1 + \frac{\theta}{\lambda} + K_0(0)\overline{R}^*(\lambda+\theta) + K_0(1)\frac{V(1)}{\theta} + K_0(2)\overline{R}^*(\lambda) + K_0(3)\frac{A'(1)}{\lambda}\Big]^{-1}.$$

Case 2: $q = 1$

Define

$$\begin{aligned}
K_1(0) &= \frac{\lambda\Big(B'(1) + C'(1)\Big) + \theta A'(1)}{N_1'(1)}, \\
K_1(1) &= \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)K_1(0) + \lambda, \\
K_1(2) &= \theta\overline{R}^*(\lambda+\theta)K_1(0), \\
K_1(3) &= V(1)K_1(1) + K_1(2) + \theta.
\end{aligned}$$

$P_{0,0}$ is given by

$$P_{0,0} = \Big[1 + \frac{\theta}{\lambda} + K_1(0)\overline{R}^*(\lambda+\theta) + K_1(1)\frac{V(1)}{\theta} + K_1(2)\overline{R}^*(\lambda) + K_1(3)\frac{A'(1)}{\lambda}\Big]^{-1}.$$

Case 3: $0 < q < 1$

Define

$$K_q(0) = \frac{\lambda B'(1)q + \Big(\lambda B(1) - \lambda - \theta\Big)g'(1) + \Big(\lambda C'(1) + \theta A'(1)\Big)q + \Big(\lambda C(1) + \theta\Big)q + \Big(g'(1) - q\Big)L(\gamma)}{N'(1)},$$

$$K_q(1) = \Big(\widetilde{R}(\lambda+\theta) + \lambda\overline{R}^*(\lambda+\theta)\Big)K_q(0) + \lambda,$$

$$K_q(2) = \frac{\theta\overline{R}^*(\lambda+\theta)K_q(0) + \overline{q}C(1)K_q(1) - \overline{q}L(\gamma) + \overline{q}\theta}{q},$$

$$K_q(3) = V(1)K_q(1) + K_q(2) + \theta.$$

$P_{0,0}$ is given by

$$P_{0,0} = \Big[1 + \frac{\theta}{\lambda} + K_q(0)\overline{R}^*(\lambda+\theta) + K_q(1)\frac{V(1)}{\theta} + K_q(2)\overline{R}^*(\lambda) + K_q(3)\frac{A'(1)}{\lambda}\Big]^{-1}.$$

Acknowledgements. This work is supported by the National Natural Science Foundation of China (No.11301306), National Science Foundation of the Anhui High Institutions of China (No.KJ2017A340), Fuyang Municipal Government-Fuyang Normal College Horizontal Cooperation projects in 2017 (No.XDHXTD201709).

References

[1] A. Aissani, S. Taleb, T. Kernane, G. Saidi and D. Hamadouche, An M/G/1 retrial queue with working vacation. In: *Proceedings of the International Conference on Systems Science 2013*, edited by J. Swiatek. Springer International Publishing, Poland (2013) 443–452.

[2] B.K. Kumar and D. Arivudainambi, The M/G/1 retrial queue with Bernoulli schedules and general retrial times. *Comput. Math. Appl.* **43** (2002) 15–30.

[3] B.K. Kumar, D. Arivudainambi and A. Vijayakumar, On the M^X/G/1 retrial queue with Bernoulli schedules and general retrial times. *Asia Pac. J. Oper. Res.* **19** (2002) 177–194.

[4] D. Arivudainambi, P. Godhandaraman and P. Rajadurai, Performance analysis of a single server retrial queue with working vacation. *OPSEARCH* **51** (2014) 434–462.

[5] G. Choudhury and J. Ke, An unreliable retrial queue with delaying repair and general retrial times under Bernoulli vacation schedule. *Appl. Math. Comput.* **230** (2014) 436–450.

[6] G.I. Falin, An M/G/1 retrial queue with an unreliable server and general repair times. *Perform. Eval.* **67** (2010) 569–582.

[7] G.I. Falin and J.G.C. Templeton, Retrial Queues. Chapman & Hall, London (1997).

[8] J. Keilson and L.D. Servi, Oscillating random walk models for GI/G/1 vacation system with Bernoulli schedules. *J. Appl. Prob.* **23** (1986) 790–802.

[9] J. Li and N. Tian, Performance analysis of a GI/M/1 queue with single working vacation. *Appl. Math. Comput.* **217** (2011) 4960–4971.

[10] J.R. Artalejo, Accessible bibliography on retrial queues: Progress in 2000–2009. *Math. Comput. Model.* **51** (2010) 1071–1081.

[11] J.R. Artalejo and A. Corral, Retrial Queueing Systems: A Computational Approach. Springer, Berlin (2008).

[12] J. Wang, Discrte-time Geo/G/1 retrial queues with general retrial time and Bernoulli vacation. *J. Syst. Sci. Complex.* **25** (2012) 504–513.

[13] J. Wu and Z. Lian, A single-server retrial G-queue with priority and unreliable server under Bernoulli vacation schedule. *Comput. Ind. Eng.* **64** (2013) 84–93.

[14] K. Chae, D. Lim and W. Yang, The GI/M/1 queue and the GI/Geo/1 queue both with single working vacation. *Perform. Eval.* **66** (2009) 356–367.

[15] L.D. Servi and S.G. Finn, M/M/1 queue with working vacations (M/M/1/WV). *Perform. Eval.* **50** (2002) 41–52.

[16] L. Tao, Z. Liu and Z. Wang, M/M/1 retrial queue with collisions and working vacation interruption under N-policy. *RAIRO: OR* **46** (2012) 355–371.

[17] M.F. Neuts, Structured Stochastic Matrices of M/G/1 Type and Their Applications. Marcel Dekker, New York (1989).

[18] M. Zhang and Z. Hou, M/G/1 queue with single working vacation. *J. Appl. Math. Comput.* **39** (2012) 221–234.

[19] N. Selvaraju and C. Goswami, Impatient customers in an M/M/1 queue with single and multiple working vacations. *Comput. Ind. Eng.* **65** (2013) 207–215.

[20] N. Tian, X. Zhao and K. Wang, The M/M/1 queue with single working vacation. *Int. J. Infor. Manag. Sci.* **19** (2008) 621–634.

[21] R.B. Cooper, Introduction to Queueing Theory. North-Holland, New York (1981).

[22] R.W. Wolff, Poisson arrivals see time averages. *Oper. Res.* **30** (1982) 223–231.

[23] S. Gao and J. Wang, Discrete-time Geo^X/G/1 retrial queue with general retrial times, working vacations and vacation interruption. *Qual. Technol. Quant. M.* **10** (2013) 493–510.

[24] S. Gao, J. Wang and W. Li, An M/G/1 retrial queue with general retrial times, working vacations and vacation interruption. *Asia Pac. J. Oper. Res.* **31** (2014) 25.

[25] S. Gao and Z. Liu, An M/G/1 queue with single working vacation and vacation interruption under Bernoulli schedule. *Appl. Math. Model.* **37** (2013) 1564–1579.

[26] S. Upadhyaya, Working vacation policy for a discrete-time Geo^X/Geo/1 retrial queue. *OPSEARCH* **52** (2015) 650–669.

[27] T. Do, M/M/1 retrial queue with working vacations. *Acta Inform.* **47** (2010) 67–75.

[28] V. Jailaxmi, R. Arumuganathan and M.S. Kumar, Performance analysis of single server non-Markovian retrial queue with working vacation and constant retrial policy. *RAIRO: OR* **48** (2014) 381–398.

[29] W. Zhou, Analysis of a single-server retrial queue with FCFS orbit and Bernoulli vacation. *Appl. Math. Comput.* **161** (2005) 353–364.