# An M/M/1 Retrial Queue with Working Vacation, Orbit Search and Balking

Juntong Li and Tao Li

*Abstract*—**In this paper, an M/M/1 retrial queue with working vacation, orbit search and balking is considered. Using the matrix-analytic method, we obtain the necessary and sufficient condition for system to be stable. We also derive the stationary probability distribution and some performance measures. Then we give the conditional stochastic decomposition for the queue length in the orbit when the server is busy. Finally, we show the effect of the model parameters on the system's characteristics by some numerical examples.**

*Index Terms*—**Retrial, Working vacation, Orbit search, Balking.**

## I. Introduction

Queueing models with server vacation have been well studied in recent years and successfully applied in every territory of life. Some vacation queues can be found in Tian and Zhang [1]. On the basis of ordinary vacation, Servi and Finn [2] first introduced a new policy as working vacation. In order to utilize the server effectively, Do [3] discussed an M/M/1 retrial queue with working vacations. Now the combination of retrial and working vacation have been studied by many scholars. Some specific results can be referenced by Li et al. [4]. On the other hand, orbit search is also innovated. Krishnamoorthy and Joshua [5] analyzed a multi-server retrial queue with search of customers from the orbit. Gao and Wang [6] considered an M/G/1 retrial queue with orbit search. Retrial queue with orbit search can also be found in [7,8]. Meanwhile, balking has also been well studied, such as [9]. Moreover, in some retrial queues, balking has been also investigated extensively. Arivudainambi and Godhandaraman [10] investigated a retrial queueing system with balking, optional service and vacation. Some other results about a retrial queue with balking can be found in [11,12,13].

To the authors' best knowledge, there is no research work investigating an M/M/1 retrial queue with working vacation, orbit search and balking. This motivates us to deal with such a queueing model in this paper. And if we let parameters in this paper take proper values, many M/M/1 queues will be special cases of our model.

The model we considered has a potential practical application in the telephone consultation of after-sale service system. Nowadays, many merchants have offered the telephone consultation services to the guests (called customers). Therefore, we construct a telephone consultation service system staffed with a main server and an assistant server. The assistant server can only provides service to the guests when the main server is rest from work, and the service rate of the assistant server is usually slower than the main server. In normal circumstances, there is a phone operator who is responsible for establishing communications between the merchants and the guests or writing down the order of the calls (corresponding to the "orbit"). When a guest makes a call, if the line is busy , the guest cannot queue but tries again sometime later (called retrial) or cancel his call and does not accept service (called balking). If the line is free, he will be served immediately by the main server or the assistant server. The guest in the orbit does not need make a call again. When a service is completed, the server may search for the guest in orbit in sequence with some probability to provide service right away (called orbit search) or with a certain probability remain idle. When the main sever can not find any guest calls, he will have a rest, and during this period, the assistant server will serve the guests (called working vacation). If there still having guests in the system when a working vacation ends, the main server will come back to provide service (start a new busy period). If there is no guest, the assistant server will continue to replace the main server.

This paper is organized as follows. In Section 2, we establish the model and derive the infinitesimal generator. In Section 3, the stationary probability distribution is obtained. In Section 4, we give a conditional stochastic decomposition. In Section 5, some numerical examples are presented to illustrate the effect of some parameters on the system's characterisitics. Finally, Section 6 concludes this paper.

## II. Model Formulation

We consider an M/M/1 retrial queue with working vacation, orbit search and balking. The detailed description of this model is given as follows:

**(1).** The interarrival times of customers are exponentially distributed with parameter $\lambda$. Upon the arrival of customers, if the server is free, service begins immediately. If the server is busy, on the other hand, customers will go to the orbit with probability $h$ $(0 \leq h \leq 1)$ or leave the system with probability $\bar{h}(\bar{h} = 1 - h)$. Request retrials from the orbit follow a Poisson process with rate $\alpha$.

**(2).** The server begins a working vacation each time when the system becomes empty, and the vacation time follows an exponential distribution with parameter $\theta$. Moreover, when a working vacation ends, if the system is non-empty at that moment, a new busy period starts. If the system is still empty, on the other hand, the server will start another working vacation. The service in a regular busy period is governed by an exponential distribution with parameter $\mu$, and in working vacation period follows an exponential distribution with parameter $\eta$.

**(3).** At the completion of a service, the server searches for the customer in orbit (if any) with probability $p$ $(0 \leq p \leq 1)$

Manuscript received; revised. This work was supported by the National Natural Science Foundation of China 11301306.

Juntong Li is with the School of Mathematics and Statistics, Shandong University of Technology.

Tao Li is with the School of Mathematics and Statistics, Shandong University of Technology, Zibo, 255049, China e-mail: liltaot@126.com.

(Advance online publication: 1 February 2019)

or remains idle with complementary probability $\bar{p}(\bar{p}=1-p)$. We assume the search time is negligible.

The interarrival times, interretrial times, service times and vacation times are also assumed to be mutually independent.

Let $Q(t)$ be the number of customers in the orbit at time $t$, and let $J(t)$ be the state of server at time $t$. There are four possible states of the server as follows:

$$J(t)=\begin{cases} 1, \textit{the server is in a working vacation} \\ \quad \textit{period at time t and the server is busy}, \\ 2, \textit{the server is in a working vacation} \\ \quad \textit{period at time t and the server is free}, \\ 3, \textit{the server is during a normal service} \\ \quad \textit{period at time t and the server is busy}, \\ 4, \textit{the server is during a normal service} \\ \quad \textit{period at time t and the server is free}. \end{cases}$$

Obviously, $\{Q(t),J(t)\}$ is a Markov process with state space,

$$\Omega=\{(0,j),j=1,2,3\}\bigcup\{(n,j),n\geq 1,j=1,2,3,4\}.$$

Using the lexicographical sequence for the states, the infinitesimal generator can be written as

$$\widetilde{Q}=\begin{pmatrix} B_1 & A_0 & & & \\ A_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix},$$

where

$$B_1=\begin{pmatrix} \beta_1 & \eta & \theta & 0 \\ \lambda & -\lambda & 0 & 0 \\ 0 & \mu & \beta_2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A_0=\begin{pmatrix} h\lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & h\lambda & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A_2=\begin{pmatrix} p\eta & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 \\ 0 & 0 & p\mu & 0 \\ 0 & 0 & \alpha & 0 \end{pmatrix},$$

$$A_1=\begin{pmatrix} \beta_1 & \bar{p}\eta & \theta & 0 \\ \lambda & \beta_3 & 0 & \theta \\ 0 & 0 & \beta_2 & \bar{p}\mu \\ 0 & 0 & \lambda & -(\lambda+\alpha) \end{pmatrix},$$

and

$$\begin{aligned} \beta_1 &= -(h\lambda+\eta+\theta), \\ \beta_2 &= -(h\lambda+\mu), \\ \beta_3 &= -(\lambda+\alpha+\theta). \end{aligned}$$

Due to the block structure of matrix $\tilde{Q}$, $\{Q(t),J(t)\}$ is called a quasi birth and death (QBD) process.

## III. Stability Condition and Stationary Distribution

*Theorem 1:* The QBD process $\{Q(t),J(t)\}$ is positive recurrent if and only if $\mu(\lambda p+\alpha)>h\lambda(\lambda+\alpha)$.

Proof: First, we assume

$$\begin{aligned} A &= A_0+A_1+A_2 \\ &= \begin{pmatrix} (p-1)\eta-\theta & \bar{p}\eta & \theta & 0 \\ \alpha+\lambda & \beta_3 & 0 & \theta \\ 0 & 0 & (p-1)\mu & \bar{p}\mu \\ 0 & 0 & \alpha+\lambda & -(\lambda+\alpha) \end{pmatrix}. \end{aligned}$$

Since matrix A is reducible, the Theorem 7.3.1 in [14] gives the condition for positive recurrence of the QBD. After permutation of rows and columns, the Theorem 7.3.1 states that the QBD is positive recurrent if and only if

$$\pi^*\begin{pmatrix} p\mu & 0 \\ \alpha & 0 \end{pmatrix}e>\pi^*\begin{pmatrix} h\lambda & 0 \\ 0 & 0 \end{pmatrix}e,$$

where $e$ is a column vector with all elements equal to one, and $\pi^*$ is the unique solution of the system $\pi^*\begin{pmatrix} (p-1)\mu & \bar{p}\mu \\ \alpha+\lambda & -(\lambda+\alpha) \end{pmatrix}=\mathbf{0}$, $\pi^*e$=1. After some algebraic manipulation, the QBD process is positive recurrent if and only if $p\mu+\frac{\alpha(\mu-p\mu)}{\lambda+\alpha}>h\lambda$, ie. $\mu(\lambda p+\alpha)>h\lambda(\lambda+\alpha)$.□

*Theorem 2:* If $\mu(\lambda p+\alpha)>h\lambda(\lambda+\alpha)$, the matrix equation $R^2A_2+RA_1+A_0=\mathbf{0}$ has the minimal non-negative solution

$$R=\begin{pmatrix} r_1 & r_2 & r_3 & r_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & r_5 & r_6 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where

$$\begin{aligned} r_1 &= \frac{[\beta_1\beta_3+\bar{p}\eta\lambda]-\sqrt{\Delta}}{2(p\eta\lambda+p\eta\theta+\eta\alpha)}, \\ r_2 &= \frac{\bar{p}\eta}{\lambda+\alpha+\theta}r_1, \\ r_3 &= \frac{r_1r_2\theta\alpha+r_1\theta(\lambda+\alpha)+r_2\theta\lambda}{-(\lambda+\alpha)[(r_1+r_5)p\mu+r_6\alpha-h\lambda-\mu]-r_1\alpha\bar{p}\mu-\bar{p}\mu\lambda}, \\ r_4 &= \frac{r_2\theta+r_3\bar{p}\mu}{\lambda+\alpha}, \\ r_5 &= \frac{h\lambda(\lambda+\alpha)}{p\mu\lambda+\mu\alpha}, \\ r_6 &= \frac{h\lambda\bar{p}\mu}{p\mu\lambda+\mu\alpha}, \end{aligned}$$

and

$$j=p\eta\lambda+p\eta\theta+\eta\alpha,$$

$$\Delta=[(h\lambda+\theta+\eta)(\lambda+\alpha+\theta)+\bar{p}\eta\lambda]^2-4jh\lambda(\lambda+\alpha+\theta).$$

Proof: From the structure of $A_0$, $A_1$ and $A_2$, we can assume $R=\begin{pmatrix} R_{11} & R_{12} \\ \mathbf{0} & R_{22} \end{pmatrix}$, where $R_{11}$, $R_{12}$ and $R_{22}$ are all $2\times 2$ matrices. Taking $R$ into $R^2A_2+RA_1+A_0=\mathbf{0}$, we have

$$\begin{cases} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}={R_{11}}^2\begin{pmatrix} p\eta & 0 \\ \alpha & 0 \end{pmatrix}+R_{11}\begin{pmatrix} \beta_1 & \bar{p}\eta \\ \lambda & \beta_3 \end{pmatrix} \\ \qquad +\begin{pmatrix} h\lambda & 0 \\ 0 & 0 \end{pmatrix}, \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}=(R_{11}R_{12}+R_{12}R_{22})\begin{pmatrix} p\mu & 0 \\ \alpha & 0 \end{pmatrix} \\ \qquad +R_{11}\begin{pmatrix} \theta & 0 \\ 0 & \theta \end{pmatrix}+R_{12}\begin{pmatrix} \beta_2 & \bar{p}\mu \\ \lambda & -(\lambda+\alpha) \end{pmatrix}, \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}={R_{22}}^2\begin{pmatrix} p\mu & 0 \\ \alpha & 0 \end{pmatrix} \\ \qquad +R_{22}\begin{pmatrix} \beta_2 & \bar{p}\mu \\ \lambda & -(\lambda+\alpha) \end{pmatrix}+\begin{pmatrix} h\lambda & 0 \\ 0 & 0 \end{pmatrix}. \end{cases}$$

From the first equation, we have $R_{11}=\begin{pmatrix} r_1 & r_2 \\ 0 & 0 \end{pmatrix}$. Similarly, $R_{22}=\begin{pmatrix} r_5 & r_6 \\ 0 & 0 \end{pmatrix}$ can be obtained from the third equation. Taking $R_{11}$ and $R_{22}$ into the second equation, we finally derived $R_{12}=\begin{pmatrix} r_3 & r_4 \\ 0 & 0 \end{pmatrix}$ by some computation. □

Under the stability condition, let $(Q,J)$ be the stationary limit of the process $\{Q(t),J(t)\}$, and denote

$$\begin{aligned}\pi_n &= (\pi_{n1},\pi_{n2},\pi_{n3},\pi_{n4}), n\geq 0,\\ \pi_{nj} &= P\{Q=n,J=j\}\\ &= \lim_{t\to\infty} P\{Q(t)=n,J(t)=j\},(n,j)\in\Omega.\end{aligned}$$

Note that when there is no customer in the orbit, the probability that the server is in a busy period and does not serve a customer is zero. Thus, $\pi_{04}=0$.

*Theorem 3:* If $\mu(\lambda p+\alpha)>h\lambda(\lambda+\alpha)$, the stationary probability distribution of $(Q,J)$ is given by

$$\begin{cases}\pi_{n1}=\pi_{01}r_1^n, \quad n\geq 1,\\ \pi_{n2}=\pi_{01}r_1^{n-1}r_2, \quad n\geq 1,\\ \pi_{n3}=\pi_{01}\left[\frac{r_3}{r_5-r_1}(r_5^n-r_1^n)\right]\\ \qquad +\pi_{03}r_5^n, \quad n\geq 1,\\ \pi_{n4}=\pi_{01}\left[r_4r_1^{n-1}+\frac{r_3r_6}{r_5-r_1}(r_5^{n-1}-r_1^{n-1})\right]\\ \qquad +\pi_{03}r_5^{n-1}r_6\,, \quad n\geq 1,\end{cases} \tag{1}$$

and

$$\begin{cases}\pi_{02}=\dfrac{(h\lambda+\eta+\theta)+r_1p\eta+r_2\alpha}{\lambda}\pi_{01},\\ \pi_{03}=\dfrac{h\lambda+\theta+r_1p\eta+r_2\alpha}{\mu}\pi_{01},\end{cases} \tag{2}$$

where $\pi_{01}$ can be determined by the normalization condition.

Proof: Using the matrix-geometric solution method, we have

$$\begin{aligned}\pi_n &= (\pi_{n1},\pi_{n2},\pi_{n3},\pi_{n4})=\pi_0R^n\\ &= (\pi_{01},\pi_{02},\pi_{03},\pi_{04})R^n, n\geq 0\end{aligned}.$$

Since

$$R^n=\begin{pmatrix} r_1^n & r_1^{n-1}r_2 & \delta_1^n & \delta_2^n\\ 0&0&0&0\\ 0&0&r_5^n&r_5^{n-1}r_6\\ 0&0&0&0\end{pmatrix}, n\geq 1,$$

where

$$\begin{aligned}\delta_1^n &= \tfrac{r_3}{r_5-r_1}(r_5^n-r_1^n),\\ \delta_2^n &= r_4r_1^{n-1}+\tfrac{r_3r_6}{r_5-r_1}(r_5^{n-1}-r_1^{n-1}),\end{aligned}$$

substituting $R^n$ into the above equation, we can obtain (1). Moreover, $\pi_0$ satisfies the next equation

$$\pi_0(B_1+RA_2)=\mathbf{0}. \tag{3}$$

Using Equation (3), we can get (2) by some caculations. Since

$$\sum_{j=0}^{4}\sum_{n=0}^{\infty}\pi_{nj}=1,$$

we can get

$$\pi_{01}=(1+x+y+z)^{-1},$$

where

$$\begin{aligned}x &= \{(1-r_1)(1-r_5)[\mu(h\lambda+\eta+\theta+r_1p\eta+r_2\alpha)\\ &\quad+\lambda(h\lambda+\theta+r_1p\eta+r_2\alpha)]\\ &\quad+\lambda\mu[(1+r_5)(r_1+r_2+r_4)\\ &\quad+(1-r_5)(r_5+r_6)]\}/[\lambda\mu(1-r_1)(1-r_5)],\\ y &= [r_3r_6\mu+r_3r_5(h\lambda+\theta+r_1p\eta\\ &\quad+r_2\alpha)]/[\mu(r_5-r_1)(1-r_5)],\\ z &= [r_3r_1(h\lambda+\theta+r_1p\eta+r_2\alpha)\\ &\quad+r_3r_6\mu]/[(r_1-r_5)(1-r_1)\mu].\end{aligned}$$ □

Clearly, the state probability of the server is given by

$$\begin{aligned}P_1 &= P\{J=1\}=\sum_{n=0}^{\infty}\pi_{n1}=\tfrac{1}{1-r_1}\pi_{01}\ ,\\ P_2 &= P\{J=2\}=\sum_{n=0}^{\infty}\pi_{n2}\\ &= \delta_3\pi_{01}+\tfrac{r_6}{(1-r_5)r_5}\delta_4\pi_{01},\\ P_3 &= P\{J=3\}=\sum_{n=0}^{\infty}\pi_{n3}\\ &= \tfrac{r_3}{r_5-r_1}\tfrac{1}{1-r_5}\pi_{01}-\tfrac{r_3}{r_5-r_1}\tfrac{1}{1-r_1}\pi_{01}+\tfrac{1}{1-r_5}\delta_4\pi_{01},\\ P_4 &= P\{J=4\}=\sum_{n=1}^{\infty}\pi_{n4}\\ &= \tfrac{r_4}{(1-r_1)r_1}\pi_{01}+\tfrac{r_6}{(1-r_5)r_5}\delta_4\pi_{01}\\ &\quad+\tfrac{r_3r_6}{r_5-r_1}\left[\tfrac{1}{(1-r_5)r_5}-\tfrac{1}{(1-r_1)r_1}\right]\pi_{01},\end{aligned}$$

where

$$\begin{aligned}\delta_3 &= \frac{r_4}{(1-r_1)r_1}+\frac{r_3r_6}{r_5-r_1}\left[\frac{1}{(1-r_5)r_5}-\frac{1}{(1-r_1)r_1}\right],\\ \delta_4 &= \frac{h\lambda+\theta+r_1p\eta+r_2\alpha}{\mu}.\end{aligned}$$

The probability that the server is busy is

$$P_b=P\{J=1\}+P\{J=3\}=P_1+P_3.$$

The probability that the server is free is

$$P_c=P\{J=2\}+P\{J=4\}=P_2+P_4=1-P_b.$$

Let $L$ be the number of customers in the orbit, we can get

$$\begin{aligned}E[L] &= \sum_{n=1}^{\infty}n(\pi_{n1}+\pi_{n2}+\pi_{n3}+\pi_{n4})\\ &= \pi_{01}\tfrac{(r_1+r_2+r_4)(1-r_5)^2+(1-r_1r_5)r_3+(2-r_1-r_5)r_3r_6}{(1-r_5)^2(1-r_1)^2}\\ &\quad+\pi_{03}\tfrac{r_5+r_6}{(1-r_5)^2}.\end{aligned}$$

Let $\widetilde{L}$ be the number of customers in the system, we have

$$\begin{aligned}E\left[\tilde{L}\right] &= \sum_{n=1}^{\infty}n(\pi_{n2}+\pi_{n4})+\sum_{n=0}^{\infty}(n+1)(\pi_{n1}+\pi_{n3})\\ &= E[L]+P_1+P_3=E[L]+P_b.\end{aligned}$$

## IV. Conditional Stochastic Decomposition

*Lemma 1:* If $\mu(\lambda p+\alpha)>h\lambda(\lambda+\alpha)$, let $Q_0$ be the conditional queue length of the retrial M/M/1 queue with orbit search and balking in the orbit given that the server is busy, then $Q_0$ has a probability generating function

$$G_{Q_0}(z)=\frac{1-r_5}{1-r_5z}.$$

Proof: The proof of this lemma is similar to the proof of Lemma 1 in Li et al. [15], so we omit it here.

Let $Q_b$ be the condition queue length of our system in the orbit when the server is busy, if $\mu(\lambda p+\alpha)>h\lambda(\lambda+\alpha)$, we

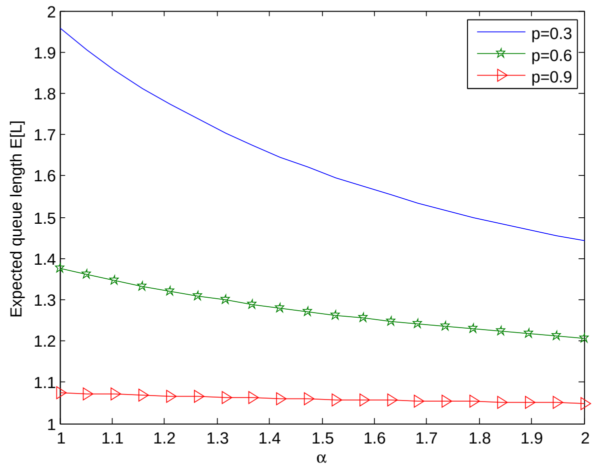


Fig. 1. The expected queue length in the orbit with the change of $\alpha$.

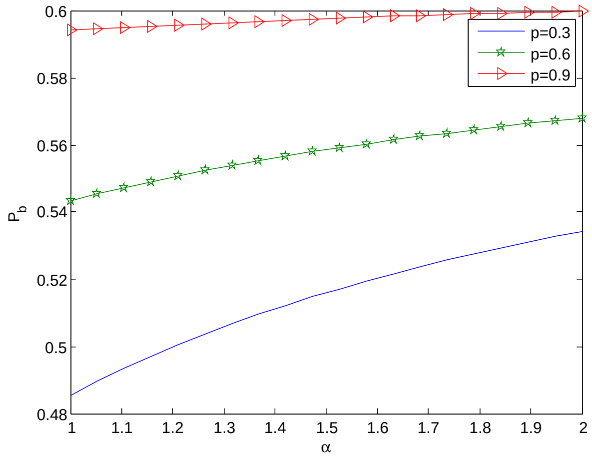


Fig. 2. The probability that the server is busy with the change of $\alpha$.

can get

$$\begin{aligned} G_{Q_b}(z) &= \sum_{n=0}^{\infty} P\{Q_b = n\} \cdot z^n \\ &= \sum_{n=0}^{\infty} \frac{\pi_{n1}+\pi_{n3}}{P_b} \cdot z^n \\ &= \frac{1-r_5}{1-r_5 z} \cdot \frac{1-r_5 z+r_3 z+\delta_4(1-r_1 z)}{(1-r_1 z)(1-r_5)P_b} \\ &= G_{Q_0}(z) G_{Q_c}(z). \end{aligned}$$

Thus, $Q_b$ can be decomposed into the sum of two independent random variables: $Q_b$=$Q_0$+$Q_c$, where $Q_0$ is defined in Lemma 1, and follows a geometric distribution with parameter $1 - r_5$, the additional queue length $Q_c$ has a probability generating function $G_{Q_c}(z)$.

*Remark 1:*

(1) If $\alpha \to \infty$, $h$=1, the model becomes an M/M/1 queue with working vacation.

(2) If $h$=1, the model reduces to an M/M/1 retrial queue with working vacation and orbit search.

(3) If $p$=0, the model becomes an M/M/1 retrial queue with working vacation and balking.

(4) If $h$=1, $p$=0, the model reduces to an M/M/1 retrial queue with working vacation.

## V. Numerical Results

In this section, under the stationary condition, we present some numerical examples to illustrate the effect of some parameters on the expected queue length $E[L]$ and the probability and the server is $P_b$. The various parameters of this model are arbitrarily chosen as $\lambda = 0.8$. $\theta = 0.3$, $\eta = 0.2$, $\mu = 1.6$, $\alpha = 0.8$, $h = 0.5$, unless they are considered as variable or their values are mentioned in the respective figures.

### A. Sensitivity Analysis

Fig.1 and Fig. 2 show the effect of $\alpha$ on the expected queue length $E[L]$ and the probability that the server is busy $P_b$, respectively. We can see that $E[L]$ is decreasing with an increasing value of $\alpha$, but $P_b$ is increasing. As the value of $\alpha$ increases, the mean retrial time decreases. From the instant when the server is free, an arriving customer and retrial customers compete to access the server. So the smaller the mean retrial time is, the bigger the probability that the server is busy is, which increases $P_b$, and decreases the value of $E[L]$. When $p = 0.9$, we can see that the effect of $\alpha$ on $E[L]$ and $P_b$ is not obvious, the reason is that the server will search for the customer in orbit with probability $p$ when a service is completed. We can also find that $E[L]$ decreases as the values of $p$ increase. While $P_b$ has an opposite tendency.

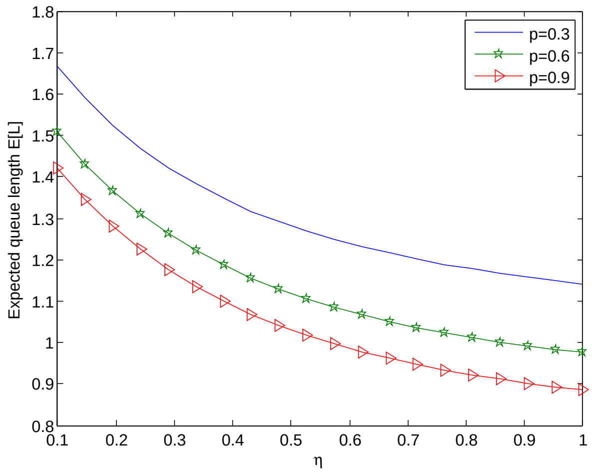


Fig. 3. The expected queue length in the orbit with the change of $\eta$.

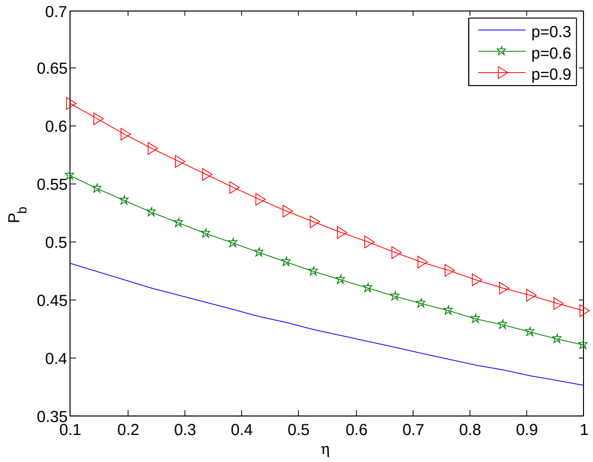


Fig. 4. The probability that the server is busy with the change of $\eta$.

In Fig.3 and Fig.4, with the change of working vacation

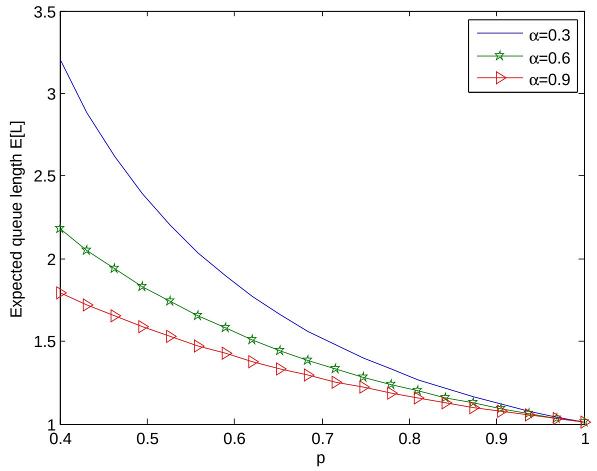


Fig. 5. The expected queue length in the orbit with the change of $p$.

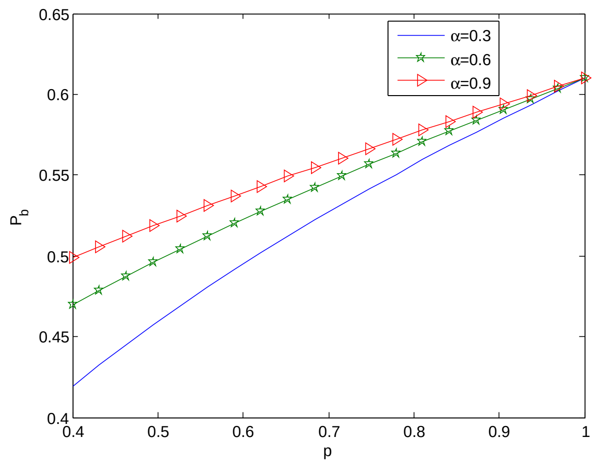


Fig. 6. The probability that the server is busy with the change of $p$.

service rate $\eta$, the curves of the expected queue length $E[L]$ and the probability $P_b$ are provided. We can find that the queue length $E[L]$ and the probability $P_b$ both decrease with an increasing value of $\eta$. This is because that the server provides service at a lower speed during the vacation period rather than stopping service completely. As a result, $E[L]$ and $P_b$ both show a downward trend.

Fig.5 and Fig.6 illustrate the effect of $p$ on the $E[L]$ and $P_b$, respectively. It is obvious that the greater the value of $p$ is, the smaller the mean orbit size is, and the bigger the probability $P_b$ is. This is due to the fact that with the value of $p$ increases, more customers can be searched from the orbit. We can also see that $E[L]$ decreases as the value of $\alpha$ increases, while $P_b$ increases with an increasing value of $\alpha$. Moreover, when $p$ approaches to 1, the model reduces to an M/M/1 queue with working vacation and balking but without retrial, we can see that the retrial rate $\alpha$ has no effect on $E[L]$ and $P_b$.

### B. Cost Analysis

In this subsection, we establish a cost function to search for the optimal service rate $\eta$, so as to minimize the expected operating cost function per unit time.

Define the following cost elements:

$C_L$=cost per unit time for each customer present in the orbit;

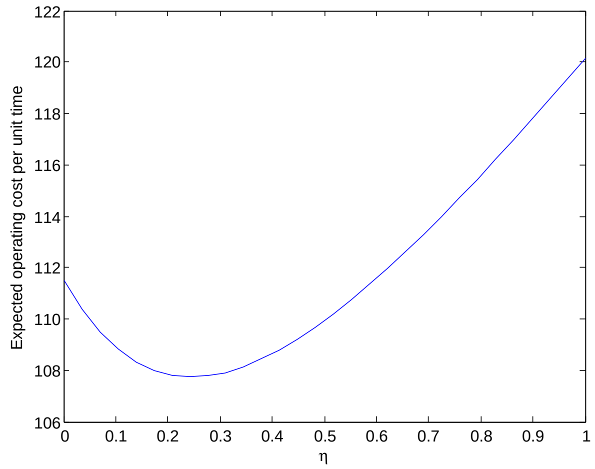


Fig. 7. The effect of $\eta$ on the expected operating cost per unit time.

$C_\mu$=cost per unit time for service during a normal service period;

$C_\eta$=cost per unit time for service in a working vacation period;

$C_\theta$=cost per unit time during a vacation period.

We establish an expected operating cost function per unit time as:

$$\min_{\eta} : f(\eta) = C_L E[L] + C_\mu \mu + C_\eta \eta + C_\theta \theta.$$

As the operating cost function per unit time is highly nonlinear and complex, it is difficult to get the derivative of it. We assume $C_L$=26, $C_\mu$=40, $C_\eta$=28, $C_\theta$=15, and use the parabolic method to find the optimum value of $\eta$, say $\eta^*$. The parabolic method generates a quadratic function through the evaluated points in each iteration, and the objective function $f(x)$ can be approximated by the quadratic function. 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

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

The parabolic method uses this approximation to improve the current 3-point pattern by replacing one of its points with an approximate optimum $\overline{x}$. Then, repeating in this way isolates an optimum for $f(x)$ in an ever-narrowing range. The step of the parabolic method can be found in [16].

TABLE I
THE PARABOLIC METHOD IN SEARCHING THE OPTIMUM SOLUTION .

| *iterations* | $x_0$ | $x_1$ | $x_2$ | $f(x_0)$ | $f(x_1)$ | $f(x_2)$ | $\overline{x}$ | $f(\overline{x})$ | *tolerance* |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.1 | 0.3 | 0.4 | 106.497033 | 105.464339 | 106.238467 | 0.260018 | 105.347749 | 0.039982 |
| 1 | 0.1 | 0.260018 | 0.3 | 106.497033 | 105.347749 | 105.464339 | 0.251132 | 105.339754 | 0.008886 |
| 2 | 0.1 | 0.251132 | 0.260018 | 106.497033 | 105.339754 | 105.347749 | 0.247162 | 105.338429 | 0.003970 |
| 3 | 0.1 | 0.247162 | 0.251132 | 106.497033 | 105.338429 | 105.339754 | 0.246076 | 105.338313 | 0.001086 |
| 4 | 0.1 | 0.246076 | 0.247162 | 106.497033 | 105.338313 | 105.338429 | 0.245640 | 105.338297 | 0.000435 |
| 5 | 0.1 | 0.245640 | 0.246076 | 106.497033 | 105.338297 | 105.338313 | 0.245508 | 105.338295 | 0.000132 |
| 6 | 0.1 | 0.245508 | 0.245640 | 106.497033 | 105.338295 | 105.338297 | 0.245459 | 105.338295 | 0.000049 |

With the information of Fig.7, we can get that there is an optimal service rate $\eta$ to make the cost minimize. Implementing the computer software MATLAB by the parabolic method and the error is controlled by $\varepsilon = 10^{-4}$. After six iterations, Table I shows that the minimum expected operating cost per unit time converges to the solution $\eta^* = 0.245459$ with a value $f(\eta^*) = 105.338295$.

## VI. Conclusion

This paper analyzes a single-server retrial queue with working vacation, orbit search and balking. Using matrix-analytic method, the condition for the system to be stable is derived, and the steady-state distributions and some performance measures are also obtained. Finally, we present some numerical examples to study the effect of various parameters, and consider a cost minimization problem. For future research, using the supplementary variable method, one can consider the similar model but with general retrial times.

## References


[1] N. Tian and Z. G. Zhang, "Vacation Queueing Models-Theory and Applications," *SpringerLink*, 2006.

[2] L. D. Servi and S. G. Finn, "M/M/1 queue with working vacations (M/M/1/WV)," *Performance Evaluation*, vol. 50, pp. 41-52, 2002.

[3] T. V. Do, "M/M/1 retrial queue with working vacations," *Acta Informatica*, vol. 47, pp. 67-75, 2010.

[4] T. Li, L. Zhang and S. Gao, "An M/M/1 retrial queue with bernoulli-schedule-controlled vacation and vacation interruption," *ICIC Express Letters*, vol. 9, no. 12, pp. 3385-3392, 2015.

[5] S. R. A. Krishnamoorthy and V. C. Joshua, "Analysis of a multi-server retrial queue with search of customers from the orbit," *Performance Evaluation*, vol. 63, no. 8, pp. 776-798, 2006.

[6] S. Gao and J. Wang, "Performance and reliability analysis of an M/G/1-G retrial queue with orbital search and non-persistent customers," *European Journal of Operational Research*, vol. 236, pp. 561-572, 2014.

[7] H. D. Meer, "Finite-source M/M/S retrial queue with search for balking and impatient customers from the orbit," *Computer Networks*, vol. 53, no. 8, pp. 1264-1273, 2009.

[8] T. G. Deepak, "On a retrial queueing model with single/batch service and search of customers from the orbit," *TOP*, vol. 23, pp. 493-520, 2015.

[9] R. Kumar and S. K. Sharma, "Queuing with reneging, balking and retention of reneged customers," *International Journal of Mathematical Models & Methods in Applied Sciences*, vol. 6, no. 7, pp. 819-828, 2012.

[10] D. Arivudainambi and P. Godhandaraman, "Retrial queueing system with balking, optional service and vacation," *Annals of Operations Research*, vol. 229, no. 1, pp. 67-84, 2015.

[11] J. C. Ke and F. M. Chang, "Modified vacation policy for M/G/1 retrial queue with balking and feedback," *Computers and Industrial Engineering*, vol. 57, pp. 433-443, 2009.

[12] A. Aboul-Hassan, S. I. Rabia and F. A. Taboly, "A discrete time Geo/G/1 retrial queue with general retrial times and balking customers," *Journal of the Korean Statistical Society*, vol. 37, no. 4, pp. 335-348, 2008.

[13] Y. Chen, C. Li and C. M. Wei, "A Discrete-Time Geo/G/1 Retrial Queue with Balking Customer, Second Optional Service, Bernoulli Vacation and General Retrial Time," *Fuzzy Systems & Operations Research and Management*, vol. 367, pp. 255-266, 2015.

[14] G. Latouche and V. Ramaswami, "Introuduction to matrix ananlytic methods in stochastic modelling," *ASA-SIAM Series on Applied Probability*, 1999.

[15] T. Li, Z. Wang and Z. Liu, "Geo/Geo/1 retrial queue with working vacations and vacation interruption," *Journal of Applied Mathematics Computing*, vol. 39, pp. 131-143, 2012.

[16] C. Luo, W. Li, K. Yu and C. Ding, " The matrix-form solution for $Geo^X$/G/1/N working vacation queue and its application to state-dependent cost control," *Computers & Operations Research*, vol. 67, pp. 63-74, 2016.