ANALYSIS OF AN $M^{(\lambda_1, \lambda_2)} / M / 1 / W V$ QUEUE WITH CONTROLLED VACATION INTERRUPTION AND VARIABLE ARRIVAL RATE

Wenqing Wu*1,2 and Yinghui Tang*1,2

1Visual Computing and Virtual Reality Key Laboratory of Sichuan Province, Sichuan Normal University, 610068, Chengdu, Sichuan, China
2School of Mathematics & Software Science, Sichuan Normal University, 610068, Chengdu, Sichuan, China
wwqing0704@163.com, tangyh@uestc.edu.cn

Abstract- This paper studies a Markovian queue with multiple working vacations and controlled vacation interruption. If there are at least N customers waiting upon completion of a service at a lower rate, the vacation is interrupted and the server returns to the system to resume the normal working level. Otherwise, the server continues the vacation until the system is non-empty after a vacation ends or there are at least N customers after a service ends. Moreover, the variable arrival rate of the customers is taken into account. Under such assumptions, by using the quasi-birth-and-death process, the matrix-geometric method and the difference equation theory, the steady-state queue length distribution along with various performance measures are derived. Additionally, under a certain cost structure, the optimal threshold $N^*$ that minimizes the long-run expected cost function per unit time is numerically determined.

Key Words- Markovian queue, working vacation, vacation interruption, cost function,

1. INTRODUCTION

Queues with server vacations have been studied extensively in the past, and have been successfully used in various applied problems. In the classical vacation queuing models, authors assumed the server is not available to serve customers during a vacation period. In 2002, Servi and Finn [1] introduced a new kind of semi-vacation policy called working vacation (WV) policy, in which the server works at a lower rate rather than completely terminates during a vacation period. Servi and Finn considered an M/M/1 queue with multiple working vacations. They derived the probability generating function of the queue length and sojourn time in steady state, and applied these results to evaluate the performance measures of gateway router in fiber communication networks. Subsequently, working vacation queues have received considerable attention in literatures,
such as Baba [2], Wu and Takagi [3], Liu et al. [4], Yu et al. [5].

In the working vacation policy, it is generally known that server resumes normal service rate only when the system is not empty at the end of a vacation. Certainly, such assumption seems much more restrictive in real world scenarios. To overcome this limitation, Li and Tian [6] introduced the vacation interruption schedule in an M/M/1 queue with working vacations. Under this type of vacation policy, if there are customers waiting at the instant of a service completion in the vacation period, the server ends his/her vacation and comes back to the normal working level; otherwise, if there is no customer waiting upon completion of the service at a lower rate, he/she continues the vacation until the system is non-empty after a service or a vacation ends. Owing to the strong application background in the stochastic service systems, many fruitful theoretical results are presented in this area. Among some excellent papers are those by Li et al. [7], Baba [8], and Zhang and Hou [9].

However, in Li and Tian’s service discipline, lower service rate only offers for the first customer arriving during a vacation period. Since the server switches from low service rate to normal service rate whenever there are customers waiting after a service completion in the vacation period, switching cost is incurred. The more the server's vacation interruption, the more additional cost the system has to face. Thus the traditional vacation interruption policy has some drawbacks in practical applications. In order to reduce the switching cost of the system, a modified vacation interruption policy is proposed in this paper. Under the modified vacation interruption policy, the server ends his/her vacation and resumes normal service rate as soon as at least N customers accumulate in the system upon the completion of a service in the vacation period. Otherwise, the server continues the vacation until the system is non-empty after a vacation ends or there are at least N customers waiting after a service ends. Meanwhile, the customers' arrival rate varies according to the status of the server. In this paper, following the construction of the long-run expected cost function per unit time, the optimal threshold N that minimizes the cost function is numerically found.

This paper is organized as follows. The next section describes the mathematical model. In Section 3, the steady-state analysis is presented and the stationary queue length distribution is derived. Some performance measures are obtained in Section 4. In Section 5, the optimal threshold value of N is found. Section 6 gives conclusions.

2. MODEL DESCRIPTION

We consider an M/M/1 queue with multiple working vacations, controlled vacation interruption and variable arrive rate, in which the server offers service at a lower rate rather than completely terminates the service during his/her vacation period. For this
system, the inter-arrival times of customers in a regular busy period are mutually independent and identically exponentially distributed with parameter $\lambda_i$. The service times provided by the server are according to exponential distributions with service rate $\mu_s$. When the system becomes empty, the server leaves for a working vacation, and the vacation duration follows an exponential distribution with mean $1/\theta$. During working vacation period, customers arrive according to a Poisson process with parameter $\lambda_v$. The working vacation period is an operation period with a lower service rate, and the service time is governed by an exponential distribution with parameter $\mu_v$. Upon completion of a service in working vacation period, if the server finds at least $N$ customers waiting in the system, the working vacation is interrupted and the server comes back to the normal working level. Otherwise, he/she continues the vacation until the system is non-empty after a vacation ends or there are at least $N$ customers after a service ends. Moreover, after completing a working vacation, the server takes another working vacation finding no customer is waiting; otherwise, the server starts providing service to customers and begins a regular busy period immediately. It is supposed that the service interrupted at the end of working vacation restarts from the beginning. Various random variables are assumed to be independent of each other.

### 3. THE STEADY-STATE ANALYSIS

Denote by $L(t)$ the number of customers in the system at time $t$, $L(t) = i (i = 0, 1, ...)$.

Let $Y(t)$ be the state of the server at time $t$, and

$$Y(t) = \begin{cases} 0, & \text{the server is in a working vacation period at time } t, \\ 1, & \text{the server is in a regular busy period at time } t. \end{cases}$$

It is easy to verify that the process $\{L(t), Y(t) : t \geq 0\}$ forms a continuous time Markov chain (CTMC) with state space $\Omega = \{(i,0) : i = 0, 1, ...\} \cup \{(i,1) : i = 1, 2, ...\}$. Further, the state transition diagram of the system is shown in Figure 1.

![State transition diagram of the system](image-url)
According to the above model assumptions, the CTMC \( \{ L(t), Y(t) : t \geq 0 \} \) is a quasi-birth-and-death process (QBD). Using the lexicographical sequence for the states, the infinitesimal generator is given by

\[
Q = \begin{pmatrix}
B_{00} & B_{01} & \ldots & \ldots & B_{0N} \\
B_{10} & A_1 & A_2 & \ldots & \ldots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
B_{N0} & A_{1} & A_{2} & \ldots & A_{N} \\
\end{pmatrix},
\]

where \( B_{ij} = -\lambda_i \), \( B_{01} = (\lambda_2, 0) \), \( B_{10} = (\mu_e) \), \( A_1 = \begin{pmatrix} -(\lambda_2 + \mu_e + \theta) & \theta \\ 0 & -(-\lambda_1 + \mu_b) \end{pmatrix} \), \( A_0 = \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{pmatrix} \), \( A_i = \begin{pmatrix} \mu_e & 0 \\ 0 & \mu_b \end{pmatrix}, i = 2, 3, ..., N \), \( A_2 = \begin{pmatrix} 0 & \mu_e \\ 0 & \mu_b \end{pmatrix} \).

First, we derive the condition for the system to reach steady state. We define matrix \( A = A_0 + A_1 + A_2 \). Thus the condition for the stability is represented by the relation \( xA_0 e < xA e \) (see Neuts [10]), where \( e \) denotes a column vector with all its elements equal to one, and \( x \) is the invariant probability vector of matrix \( A \), namely, \( xA = 0 \) and \( xe = 1 \). After some manipulation, the stability condition turns out to be

\[
\rho \triangleq \frac{\lambda}{\mu_b} < 1.
\]

Let \( \pi \), partitioned as \( \pi = (\pi_{0,0}, \pi_1, \pi_2, \ldots) \) and \( \pi_i = (\pi_{i,0}, \pi_{i,1}) \), \( i = 1, 2, ... \) be the steady-state probability vector of \( Q \). We have that from the stability condition (1),

\[
\pi_k = (\pi_{k,0}, \pi_{k,1}) = \pi_0 R^{k-N} = (\pi_{N,0}, \pi_{N,1}) R^{k-N}, k \geq N,
\]

where the matrix \( R \) is the minimal nonnegative solution to the matrix-quadratic equation

\[
R^2 A_2 + RA_1 + A_0 = 0.
\]
triangular, we suppose the rate matrix $R$ has the same structure as $R = \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}$.

Substituting $R$ into Eq.(3), the explicit expression for the rate matrix $R$ is

$$R = \begin{pmatrix} r & \frac{\lambda_2 (\lambda_2 + \theta)}{\mu_b (\lambda_2 + \mu_v + \theta)} \\ 0 & \rho \end{pmatrix}$$

and

$$R^k = \begin{pmatrix} r^k & \frac{\lambda_2 (\lambda_2 + \theta)}{\mu_b (\lambda_2 + \mu_v + \theta)} \sum_{j=0}^{k-1} r^j \rho^{k-1-j} \\ 0 & \rho^k \end{pmatrix}, k = 1, 2, ..., \rho$$

where $r = \frac{\lambda_2}{\lambda_2 + \mu_v + \theta}$.

To obtain $\pi_{i,j} ((i, j) \in \Omega)$, the following boundary equations should be solved

$$\begin{align*}
\pi_0 B_{00} + \pi_1 B_{10} &= 0, \\
\pi_0 B_{01} + \pi_1 A_1 + \pi_2 B_2 &= 0, \\
\pi_{i-1} A_0 + \pi_i A_i + \pi_{i+1} B_{i+1} &= 0, i = 2, 3, ..., N - 1, \\
\pi_{N-1} A_0 + \pi_N (A_N + R A_2) &= 0.
\end{align*}$$

Then, it follows that

$$\begin{align*}
\lambda_2 \pi_{0,0} &= \mu_v \pi_{1,0} + \mu_b \pi_{1,1}, \\
(\lambda_2 + \mu_v + \theta) \pi_{1,0} &= \lambda_2 \pi_{i-1,0} + \mu_v \pi_{i+1,0}, i = 1, 2, ..., N - 1, \\
(\lambda_2 + \mu_v + \theta) \pi_{N,0} &= \lambda_2 \pi_{N-1,0}, \\
(\lambda_1 + \mu_b) \pi_{1,1} &= \mu_b \pi_{2,1} + \theta \pi_{1,0}, \\
(\lambda_1 + \mu_b) \pi_{i,1} &= \lambda_1 \pi_{i-1,1} + \mu_b \pi_{i+1,1} + \theta \pi_{i,0}, i = 2, 3, ..., N - 1, \\
\mu_b \pi_{N,1} &= \lambda_1 \pi_{N-1,1} + [r (\lambda_2 + \mu_v + \theta) + \theta] \pi_{N,0}.
\end{align*}$$

We rewrite Eq.(5) as

$$\mu_v \pi_{i,1} - (\lambda_2 + \mu_v + \theta) \pi_{i,0} + \lambda_2 \pi_{i+1,0} = 0, i = 1, 2, ..., N - 1,$$

which is a linear homogeneous difference equation of order 2 with constant coefficients and the quadratic characteristic equation $\mu_v y^2 - (\lambda_2 + \mu_v + \theta) y + \lambda_2 = 0$. Obviously, the quadratic characteristic equation has two roots, $\sigma_1$ and $\sigma_2$, given by
\[
\sigma_{1,2} = \frac{(\lambda_2 + \mu_i + \theta) \pm \sqrt{(\lambda_2 + \mu_i + \theta)^2 - 4\mu_i\lambda_2}}{2\mu_i}.
\]

Therefore, from the theory of linear homogeneous difference equation with constant coefficients (see Elaydi [11]), the general solution of Eq.(10) are

\[
\pi_{i,0} = M_1\sigma_1^i + M_2\sigma_2^i, i = 0,1,\ldots,N,
\]

where \(M_1\) and \(M_2\) are constants to be determined.

To find the constants \(M_1\) and \(M_2\), substituting Eq.(11) into Eqs.(4) and (6) gives

\[
\begin{align*}
\left[(\lambda_2 - \mu_i, \sigma_1)M_1 + (\lambda_2 - \mu_i, \sigma_2)M_2 = \mu_i\pi_{1,1},
\left[(\lambda_2 + \mu_i + \theta)\sigma_1^N - \lambda_2\sigma_1^{N+1}\right]M_1 + \left[(\lambda_2 + \mu_i + \theta)\sigma_2^N - \lambda_2\sigma_2^{N+1}\right]M_2 = 0.
\end{align*}
\]

After some algebraic manipulation, we obtain that \(M_i = \Phi_i\pi_{i,1}, i = 1,2,\ldots\), where

\[
\Phi_j = \frac{\mu_i\left[(\lambda_2 + \mu_i + \theta)\sigma_1^N - \lambda_2\sigma_1^{N+1}\right]}{\sum_{m=1}^{\infty}(-1)^{m+1}\left[(\lambda_2 - \mu_i, \sigma_m)\left[(\lambda_2 + \mu_i + \theta)\sigma_m^N - \lambda_2\sigma_m^{N+1}\right]\right]}, j = 1,2.
\]

Hence, \(\pi_{i,0}\) can be expressed in terms of \(\pi_{i,1}\) as follows

\[
\pi_{i,0} = M_1\sigma_1^i + M_2\sigma_2^i = (\Phi_1\sigma_1^i + \Phi_2\sigma_2^i)\pi_{1,1}, i = 0,1,\ldots,N.
\]

Applying Eq.(12), we rewrite Eq.(8) as

\[
\mu_i\pi_{i,1} + (\lambda_2 + \mu_i)\pi_{i,1} + \lambda_2\pi_{i-1,1} + M_1\theta\sigma_1^i + M_2\theta\sigma_2^i = 0, i = 2,3,\ldots,N-1.
\]

Eq.(13) shows that the stationary probability \(\pi_{i,i}(i = 1,2,\ldots,N)\) is the solution of the following linear nonhomogeneous difference equation

\[
\mu_i\pi_{i,1} - (\lambda_2 + \mu_i)\pi_{i,1} + \lambda_2\pi_{i-1,1} + M_1\theta\sigma_1^i + M_2\theta\sigma_2^i = 0, i = 2,3,\ldots,N.
\]

The general solution of Eq.(14), denoted by \(z_i\), has the following structure

\[
z_i = z_i^p + z_i^c, i = 1,2,\ldots,N,
\]

where \(z_i^c = C_1 + C_2\rho^i\), it represents the general solution of the associated homogeneous equation \(\mu_i\pi_{i,1} - (\lambda_2 + \mu_i)\pi_{i,1} + \lambda_2\pi_{i-1,1} = 0\), and \(z_i^p\) is a particular solution of Eq.(14).

Because of the non-homogeneous term is not a solution of the associated
homogeneous equation, we set \( z_i' = D_i \sigma_i + D_2 \sigma_2^i \). It follows from Eq.(14) by taking into account \( z_i' \) that \( D_i = \Psi_i \Phi_i \pi_{1,i}, i = 1, 2, \) where \( \Psi_i = -\frac{\theta \sigma_i}{\mu_i \sigma_i^2 - (\lambda_i + \mu_i) \sigma_i + \lambda_i}, i = 1, 2. \)

We now turn our attention to finding the values of \( C_1 \) and \( C_2 \) in general solution \( z_i \). Employing Eq.(7) and noting that \( \pi_{1,i} = z_i \), we obtain

\[
\begin{cases}
(\lambda_i + \mu_i) \pi_{1,i} = \mu_i \left( C_1 + C_2 \rho_1^2 + D_1 \sigma_1^i + D_2 \sigma_2^i \right) + \theta \left( M_1 \sigma_i + M_2 \sigma_2 \right), \\
\pi_{1,i} = C_1 + C_2 \rho + D_1 \sigma_1 + D_2 \sigma_2.
\end{cases}
\]

Solving the above equations, it finally yields \( C_i = H_i \pi_{1,i}, i = 1, 2, \) where

\[
H_{i} = \frac{\mu_i - \theta \sum_{m=1}^{2} \Phi_m \sigma_m + \sum_{m=1}^{2} \Psi_m \Phi_m \sigma_m (-\mu_m \sigma_m)}{\mu_i (1 - \rho)} , \quad H_{2} = \frac{\theta \sum_{m=1}^{2} \Phi_m \sigma_m - \mu_i \sum_{m=1}^{2} \Psi_m \Phi_m \sigma_m (1 - \sigma_m) - \lambda_i}{\mu_i \rho (1 - \rho)} .
\]

Hence, \( \pi_{1,i} (i = 1, 2, \ldots, N) \) are given as

\[
\pi_{1,i} = \left( H_1 + H_2 \rho^i + \Psi_1 \Phi_1 \sigma_1^i + \Psi_2 \Phi_2 \sigma_2^i \right) \pi_{1,i}, i = 1, 2, \ldots, N. \quad (15)
\]

Furthermore, Substituting \( \pi_{N,0}, \pi_{N,1} \) and \( R^{k-N} \) into Eq.(2), we have that

\[
\pi_{k,0} = \left( \Phi_1 \sigma_1^N + \Phi_2 \sigma_2^N \right) r^{k-N} \pi_{1,1}, k = N, N + 1, 2, \ldots, \quad (16)
\]

\[
\pi_{k,0} = \left[ \left( \Phi_1 \sigma_1^N + \Phi_2 \sigma_2^N \right) - \frac{1}{\mu_i} \frac{\lambda_i (1 + \theta + \theta)}{\mu_i (\lambda_i + \mu_i + \theta)} \right] \pi_{1,1} \right) r^{k-N-1-j} + \left( H_1 + H_2 \rho^N + \Psi_1 \Phi_1 \sigma_1^N + \Psi_2 \Phi_2 \sigma_2^N \right) \rho^{k-N} \right] \pi_{1,i}, k = N + 1, N + 2, \ldots. \quad (17)
\]

Up to now we obtain the stationary probability \( \pi_{k,0} (k = 0, 1, \ldots) , \pi_{k,1} (k = 2, 3, \ldots) \) in terms of \( \pi_{1,i} \). Using the normalization condition \( \sum_{k=1}^{\infty} \pi_{k,0} + \sum_{k=1}^{\infty} \pi_{k,1} = 1 \), which finally yields

\[
\pi_{1,i} = \Delta , \text{ where } \Delta^{-1} = \sum_{j=1}^{2} \Phi_j \frac{1}{1 - \sigma_j^i} + \frac{1}{1 - r} \sum_{j=1}^{2} \Phi_j \sigma_j^i + NH_i + \sum_{j=1}^{2} \Psi_j \Phi_j \frac{(1 - \sigma_j^N)}{1 - \sigma_j}. \]
\[ + \frac{\lambda_2 (\lambda_2 + \theta)}{(1-r)(1-\rho)} \mu_b \left( \frac{\lambda_2 + \mu_r + \theta}{(1-r)^2} \sum_{j=1}^{2} \Phi_j \sigma_j^N \right) + \frac{\rho}{1-\rho} \left( H_1 + H_2 + \sum_{j=1}^{2} \Psi_j \Phi_j \sigma_j^N \right). \]

Now, all \( \pi_{i,j} \) \( (i, j) \in \Omega \) are completely determined.

3. SYSTEM PERFORMANCE MEASURES

- The average queue length when the server is on a working vacation is
  \[ E[L_0] = \left\{ \frac{2 \sum_{j=1}^{2} \Phi_j \sigma_j}{(1-\sigma_j)^2} + \frac{r((N+1)-Nr)}{(1-r)^2} \sum_{j=1}^{2} \Phi_j \sigma_j^N \right\} \Delta. \]

- The average queue length when the server is in regular busy period is
  \[ E[L_1] = \left\{ \frac{\rho(1-(N+1)\rho^N + N\rho^{N+1})}{(1-\rho)^2} + \frac{2 \sum_{j=1}^{2} \Psi_j \Phi_j \sigma_j}{(1-\sigma_j)^2} \right\} \frac{1-(N+1)\sigma_j^N + N\sigma_j^{N+1}}{1-r} \sum_{j=1}^{2} \Phi_j \sigma_j^N \]
  \[ + \frac{N(N+1)}{2} H_1 + r^{2} \frac{(N+1)-Nr}{(1-r)^2} \left( H_1 + H_2 \rho^N + \sum_{j=1}^{2} \Psi_j \Phi_j \sigma_j^N \right) \right\} \Delta. \]

- The average queue length is given by
  \[ E[L] = E[L_0] + E[L_1]. \]

- The probability that the server is on a working vacation is given by
  \[ P_{WV} = \left\{ \frac{2 \sum_{j=1}^{2} \Phi_j \sigma_j^{N+1}}{1-\sigma_j} + \frac{r}{1-r} \sum_{j=1}^{2} \Phi_j \sigma_j^N \right\} \Delta. \]

- The probability that the server’s vacation interruption is given by
  \[ P_{IR} = \frac{r}{1-r} \left( \Phi_1 \sigma_1^N + \Phi_2 \sigma_2^N \right) \Delta. \]

5. OPTIMIZATION ANALYSIS

In this subsection, we construct the long-run expected cost function per unit time for the system under consideration, in which \( N \) is a key decision variable. Our purpose is to determine the optimal threshold, say \( N^* \), so as to minimize the cost function. Let us
define the following cost elements:

c₀ = unit time cost of every customer when the server is on a working vacation,

c₁ = unit time cost of every customer when the server is in a regular busy period,

c₂ = fixed service cost per unit time during a regular busy period,

c₃ = fixed service cost per unit time during a working vacation period,

c₄ = fixed cost every time for the server’s vacation interruption.

Utilizing the definition of each cost element listed above and its corresponding system performance measures, the long-run expected cost function per unit time is

\[
C(N) = c₀E[L₀] + c₁E[L₁] + c₂μ_b + c₃μ_v + c₄μ_v P_{\text{Interruption}}
\]  

\( (18) \)

It would be an arduous task to develop the optimal value \( N^* \) analytically owing to the highly non-linear and complex nature of the long-run expected cost function. Therefore, we will numerically determine the optimal threshold \( N^* \) for \( C(N) \).

To accomplish this, a numerical example is provided by considering the following cost elements \( c₀ = 16 \), \( c₁ = 10 \), \( c₂ = 30 \), \( c₃ = 18 \), \( c₄ = 180 \), and other system parameters are taken as \( \lambda₁ = 0.75 \), \( \lambda₂ = 0.6 \), \( ω_b = 2.5 \), \( ω_v = 0.45 \), \( θ = 0.25 \). Substituting these values into \( C(N) \), we obtain the results displayed in Table 1 and Figure 2 for different values of \( N \). From the computational results, it appears that \( C(N^*) = C(4) = 105.5546 \) is the minimum of the long-run expected cost.

Table 1. The long-run expected cost per unit time for different values of \( N \).

<table>
<thead>
<tr>
<th>N</th>
<th>C(N)</th>
<th>N</th>
<th>C(N)</th>
<th>N</th>
<th>C(N)</th>
<th>N</th>
<th>C(N)</th>
<th>N</th>
<th>C(N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>113.2735</td>
<td>7</td>
<td>106.8749</td>
<td>13</td>
<td>108.8103</td>
<td>19</td>
<td>110.1112</td>
<td>25</td>
<td>111.3559</td>
</tr>
<tr>
<td>2</td>
<td>107.4951</td>
<td>8</td>
<td>107.3105</td>
<td>14</td>
<td>109.0405</td>
<td>20</td>
<td>110.3194</td>
<td>26</td>
<td>111.5628</td>
</tr>
<tr>
<td>3</td>
<td>105.7573</td>
<td>9</td>
<td>107.6880</td>
<td>15</td>
<td>109.2623</td>
<td>21</td>
<td>110.5272</td>
<td>27</td>
<td>111.7697</td>
</tr>
<tr>
<td>4</td>
<td>105.5546</td>
<td>10</td>
<td>108.0157</td>
<td>16</td>
<td>109.4786</td>
<td>22</td>
<td>110.7346</td>
<td>28</td>
<td>111.9766</td>
</tr>
<tr>
<td>5</td>
<td>105.8971</td>
<td>11</td>
<td>108.3052</td>
<td>17</td>
<td>109.6914</td>
<td>23</td>
<td>110.9418</td>
<td>29</td>
<td>112.1835</td>
</tr>
<tr>
<td>6</td>
<td>106.3877</td>
<td>12</td>
<td>108.5673</td>
<td>18</td>
<td>109.9020</td>
<td>24</td>
<td>111.1489</td>
<td>30</td>
<td>112.3904</td>
</tr>
</tbody>
</table>
6. CONCLUSION

This paper considered an $M^{(\lambda_1, \lambda_2)}/M/1$ queueing system with working vacations, controlled vacation interruption and variable arrival rate that has potential applications in modeling the industrial systems, the manufacturing systems and others. In this model, a variety of system performance measures are derived. Moreover, the long-run expected cost function per unit time of the system is developed. Meanwhile, the optimal threshold $N^*$ and the minimum cost $C(N^*)$ are numerically found. For future research, one can consider the same model where the service times follow phase-type distributions.

7. ACKNOWLEDGMENTS

This research is supported by the project of Visual Computing and Virtual Reality Key Laboratory of Sichuan Province (No. KJ201401), and the National Natural Science Foundation of China (No.71171138).

8. REFERENCES

1. L.D. Servi and S.G. Finn, M/M/1 queues with working vacations (M/M/1/WV), Performance Evaluation 50(1), 41-52, 2002.