Performance Analysis of Retrial Queueing Model with Working Vacation, Interruption, Waiting Server, Breakdown and Repair

P. Gupta*, N. Kumar

Department of Mathematics, Baba Mastnath University, Asthal Bohar, Rohtak, India

Received 20 March 2021, accepted in final revised form 15 July 2021

Abstract

In this present paper, an M/M/1 retrial queueing model with a waiting server subject to breakdown and repair under working vacation, vacation interruption is considered. Customers are served at a slow rate during the working vacation period, and the server may undergo breakdowns from a normal busy state. The customer has to wait in orbit for the service until the server gets repaired. Steady-state solutions are obtained using the probability generating function technique. Probabilities of different server states and some other performance measures of the system are developed. The variation in mean orbit size, availability of the server, and server state probabilities are plotted for different values of breakdown parameter and repair rate with the help of MATLAB software. Finally, cost optimization of the system is also discussed, and the optimal value of the slow service rate for the model is obtained.

Keywords: Retrial queue; Working vacation; Interruption; Breakdown; Optimization.

© 2021 JSR Publications. ISSN: 2070-0237 (Print); 2070-0245 (Online). All rights reserved. doi: http://dx.doi.org/10.3329/jsr.v13i3.52546

1. Introduction

Queueing models with unreliable server and server vacations have paid attention in recent years due to their applications in every field of life. Due to wide applications in telecommunication networks and computer services, retrial queues have been an intensive research topic for a long time. During working vacations, the server does not stop working completely; instead, the server provides service at a slow rate. The unreliable servers may break down, due to which customers' service is temporarily stopped until the system gets repaired. Servi and Finn [1] first introduced the queueing model with a working vacation policy. This new vacation class enhances the system's performance as the server provides slow service in this vacation instead of completely staying idle on vacations. Xu and Tian [2] analyzed M/M/1 working vacation queues with setup times. This concept of set up time enables power saving. A pioneer work on M/G/1 queues with working vacation and vacation interruption was done by Zhang and Hou [3]. The concept of vacation interruption improves the system performance as the server returns to the normal service

* Corresponding author: poonammital2207@gmail.com
period on service completion of a customer in vacations if there are still customers waiting for the service. Li and Tian [4] analyzed queues with working vacations and interruption. M/M/1 queue with a working vacation in a Fuzzy environment was studied by Kannadasan and Sathiyamoorth [5] using supplementary variables techniques.

In retrial queueing systems, the customers who find the server busy are forced to join the orbit of infinite capacity to retry for their service request. Generally, queueing systems with repeated attempts operate under classical and constant retrial policies. In classical retrial policy, each customer performs repeated attempts independently of all the other customers in orbit, whereas in constant retrial policy, only the customer at the head of the queue in orbit can retry for the service. Ayyappan et al. [6] used the truncation method to find steady-state probabilities of a queueing system with working vacation under retrial policy and preemptive priority service. This approach models the congestion systems with priority discipline and facility of retrials for the service. Li et al. [7] studied Geo/Geo/1 retrial queues with working vacation and vacation interruption. Li et al. [8] analyzed a single server retrial queueing model with working vacation, vacation interruption. The concept of working vacation with interruption policy in retrial queueing systems reduces customers’ waiting time in orbit as the server is available during such a vacation and resumes its normal service rate on completion of service in vacation, finding at least one customer waiting for his turn.

In the queueing literature, it is generally assumed that server is permanently available for the service. However, it seems unrealistic as the server may undergo a breakdown at any point in time. Many authors [9-11] analyzed retrial queueing systems with the unreliable server. Choudhury and Ke [12] have done pioneer work on queueing systems with delaying the repair. The phenomenon of delayed repair enables the server to continue service for some time even when it is under a breakdown state before sending for repair. Choudhury et al. [13] studied the concept of unreliable server with two phases of service under general retrial times and Bernoulli vacation schedule. This approach allows the server to leave for vacations with some probability and stay in a system with complementary probability. A steady-state solution for the retrial queueing model with Bernoulli vacation and server breakdown was analyzed by Rajadurai et al. [14,15]. This work combines the two important concepts of Bernoulli schedule vacations and server unreliability. Chandrasekaran et al. [16] performed a survey on queueing model with working vacation. This survey provides excellent guidance for beginners researching this field. Janani and Lakshmi [17] studied transient state solutions of M/M/1 queues with working vacation under disaster and repair. The transient solution provides information regarding the time-dependent variations. The retrial queue with working vacations and starting failure were studied by Yang and Wu [18]. Ke et al. [19] discussed retrial queue with impatient behavior of customers subject to server breakdowns. These approaches include the impact of starting failures and impatience of customers in retrial congestion systems.

In this paper, we have extended the work of Li et al. [8] under classical retrial policy by including the concept of waiting server, server breakdown, and repair to make the
model more realistic. This extension is needed to model the real-life congestion problems more closely because a completely reliable queueing model is only an ideal concept that is not compatible with practical situations of the real world. Thus, to implement retrial queueing models, it is necessary to incorporate the breakdown and repair of the server more practically. To the best knowledge of authors, there is no work published in queueing literature with retrial queueing systems under working vacations, vacation interruption subjected to waiting server, breakdown, and repair, using probability generating functions.

2. Model Description

We consider an $M/M/1$ retrial queueing model with a waiting server, server breakdown, repair under working vacation, and interruption. The model description is explained below:

1. Customer arrives in the system with arrival rate $\lambda$ which follows Poisson distribution, i.e., the time between two consecutive arrivals is exponentially distributed with mean $1/\lambda$.
2. Customers get service on the basis of FCFS queue discipline.
3. Customers get service immediately upon arrival if the server is free. The service time is exponentially distributed, having parameter $\mu$. On the other hand, if the server is busy, customers enter into orbit (free pool) having infinite capacity. Customers reattempt for their service independent of one another with rate $\zeta$. The retrial time is again exponentially distributed.
4. When the system becomes empty, the server will immediately wait instead of going on vacation. The waiting time is exponentially distributed with parameter $\gamma$. The server provides service to the customers if he arrives before the waiting timer expires; otherwise, the server goes for a working vacation where service is provided at a slow rate $\phi$. Vacation time is exponentially distributed with parameter $\theta$. Moreover, at the end of the customer's service on vacation, if customers are present in the system, then vacation is interrupted. The server returns to the normal service period, but if the system is empty at service completion instant in working vacation state, the vacation will continue.
5. If the system is empty at the end of the vacation period, then the server goes for another vacation and so on; otherwise, will return from vacation. In addition, there may occur server breakdown during a normal busy state. Whenever the system undergoes breakdown, then the system needs repair. Time to system breakdown and repair times of the system are exponentially distributed with parameters $\alpha$ and $\beta$, respectively. After repair, the server returns to provide service to customers.
3. Balance Equations and PGFs

Let \( C(t) \) denotes the number of customers in orbit at time \( t \), and the server state is represented by \( H(t) \) at a given time \( t \). The possible values of the server states can be 1 through 5 defined as follows:

- \( H(t) = 1, 2 \) indicates that the server is in a free and in a busy (serving customers) state in a normal period.
- \( H(t) = 3, 4 \) denotes in the vacation period, the server is free and in a busy (serving customers) state.
- \( H(t) = 5 \) represents the server is under repair.

Thus \( \{C(t), H(t)\} \) is a Markov process having state space \( S = \{(n, h), n \geq 1, h=1,2,4,5\} \cup \{(0, h), h=1,2,3,4\} \). Due to vacation interruption policy and assumption that server may undergoes breakdown during normal busy period. So, states \( \{(n, 3), n \geq 1\} \) and \( (0,5) \) does not make sense.

Denoting the steady-state probabilities by \( p_{n,h} \), where \( (n, h) \in S \). the balance equations are as follows:

\[
\begin{align*}
(\lambda + \gamma)p_{0,1} &= \mu p_{0,2} \quad (1) \\
(\lambda + n\delta)p_{n,1} &= \mu p_{n,2} + \phi p_{n,4} + \beta p_{n,5}, \quad n \geq 1 \quad (2) \\
(\lambda + \mu + \alpha)p_{0,2} &= \lambda p_{0,1} + \theta p_{0,4} + \xi p_{1,1} \quad (3) \\
(\lambda + \mu + \alpha)p_{n,2} &= \lambda p_{n-1,2} + \theta p_{n,4} + (n + 1)\xi p_{n+1,1} + \lambda p_{n,1}, \quad n \geq 1 \quad (4) \\
\lambda p_{0,3} &= \gamma p_{0,1} + \phi p_{0,4} \quad (5) \\
(\lambda + \theta + \phi)p_{0,4} &= \lambda p_{0,3} \quad (6) \\
(\lambda + \theta + \phi)p_{n,4} &= \lambda p_{n-1,4}, \quad n \geq 1 \quad (7) \\
(\lambda + \beta)p_{1,5} &= \alpha p_{0,2} \quad (8) \\
(\lambda + \beta)p_{n,5} &= \alpha p_{n-1,2} + \lambda p_{n-1,5}, \quad n \geq 2 \quad (9)
\end{align*}
\]

Probability generating functions are defined as

\[
\begin{align*}
F_1(z) &= \sum_{n=0}^{\infty} p_{n,1}z^n \\
F_2(z) &= \sum_{n=1}^{\infty} p_{n,2}z^n \\
F_4(z) &= \sum_{n=0}^{\infty} p_{n,4}z^n \\
F_5(z) &= \sum_{n=1}^{\infty} p_{n,5}z^n
\end{align*}
\]

Using probability generating functions from equations (1), (2), and (3), we get

\[
\lambda F_1(z) + \xi z F_1'(z) = \mu F_2(z) + \phi F_4(z) + \beta F_5(z) - \lambda p_{0,3} \tag{10}
\]
Similarly, equations (3) and (4) yield
\[(\lambda + \mu + \alpha)F_2(z) = \lambda F_3(z) + \lambda z F_2(z) + \theta F_4(z) + \xi F_1'(z) \quad (11)\]

On the same steps, equations (5) and (6) give
\[(\lambda + \theta + \phi)F_4(z) = \lambda z F_4(z) + \lambda p_{03} \quad (12)\]

From equations (7) and (8)
\[(\lambda + \beta)F_5(z) = \alpha z F_2(z) + \lambda z F_5(z) \quad (13)\]

Using equations (11) and (13) in equation (10), we get
\[
\xi \left( z - \frac{\mu(\lambda + \beta) + z(\alpha \beta - \lambda \mu)}{(\lambda + \beta - \lambda z)(\lambda + \mu + \alpha - \lambda z)} \right) F_1'(z) \\
+ \lambda \left( 1 - \frac{\mu(\lambda + \beta) + z(\alpha \beta - \lambda \mu)}{(\lambda + \beta - \lambda z)(\lambda + \mu + \alpha - \lambda z)} \right) F_1(z) \\
= \left( \phi + \frac{\theta(\mu(\lambda + \beta) + z(\alpha \beta - \lambda \mu))}{(\lambda + \beta - \lambda z)(\lambda + \mu + \alpha - \lambda z)} \right) F_4(z) - \lambda p_{03}
\]

Which implies
\[
\xi (z g_2(z) - g_1(z)) F_1'(z) + \lambda (g_2(z) - g_1(z)) F_1(z) \\
= \left( (\theta g_1(z) + \phi g_2(z)) \frac{\lambda}{\lambda + \theta + \phi - \lambda z} - \lambda g_2(z) \right) p_{03} \quad (14)
\]

Where,
\[g_1(z) = \mu(\lambda + \beta) - \lambda \mu z + \alpha \beta z\]
\[g_2(z) = (\lambda + \beta - \lambda z)(\lambda + \mu + \alpha - \lambda z)\]

To solve the differential equation, multiply both sides of equation (14) by Integrating Factor
\[I(z) = \left( (z - b)^2 - a^2 \right)^{\frac{\lambda}{2}} \frac{\lambda^2 b - c}{2\lambda a} \]

where,
\[a = \sqrt{(\alpha + \beta + \lambda + \mu)^2 - 4 \mu(\lambda + \beta)} \]
\[b = \frac{\alpha + \beta + \mu + \lambda}{2\lambda} \]
\[c = \lambda^2 + \lambda \alpha + \lambda \beta + \alpha \beta \]

And on integrating, we get
Solving equations (10) and (11), simultaneously, we get

\[ F_2(z) = \frac{(\lambda + \beta - \lambda z)(\lambda - \lambda z)F_1(z) - (\theta z + \phi)F_4(z) + \lambda p_{03}}{g_1(z) - zg_2(z)} \]  \hspace{1cm} (16)

On taking the limit in equations (12) and (15), we get

\[ F_4(1) = \frac{\lambda}{\theta + \phi} p_{03} \]  \hspace{1cm} (17)

\[ F_1(1) = \frac{1}{I(1)} \int_0^1 I(z) \left( \frac{\lambda g_1(z) + \phi g_2(z)}{\lambda + \theta + \phi - \lambda z} - \lambda g_2(z) \right) \frac{p_{03}}{\xi z(g_2(z) - g_1(z))} dz \]  \hspace{1cm} (18)

Where,

\[ I(1) = ((1 - b)^2 - a^2)\frac{\lambda}{2} \left( 1 - \frac{b - a}{1 + b + a} \right) \]

On differentiating equation (12) and taking limits, we have

\[ F_4'(1) = \frac{\lambda^2}{(\theta + \phi)^2} p_{03} \]  \hspace{1cm} (19)

\[ F_4''(1) = \frac{2\lambda^2}{(\theta + \phi)^3} p_{03} \]  \hspace{1cm} (20)

Using the L-Hospital rule, from equation (16), we get

\[ F_2(1) = \frac{\beta(\lambda F_1(1) + \theta F_4(1) + (\theta + \phi)F_4'(1))}{\mu \beta - \lambda (\alpha + \beta)} \]  \hspace{1cm} (21)

\[ F_5(1) = \frac{a}{b} F_2(1) \]  \hspace{1cm} (22)

Taking limits on equation (11)

\[ F_1'(1) = \frac{(\mu + \alpha)F_2(1) - \lambda F_4(1) - \theta F_4'(1)}{\xi} \]  \hspace{1cm} (23)

On differentiating equation (16) and using equations (10), (13) gives

\[ F_2'(1) = \frac{\beta (\mu \beta - \lambda \beta - \lambda \alpha)((\theta + \phi)F_4'(1) + 2\lambda F_4'(1) + 2\theta F_4'(1))}{2(\alpha \lambda + \lambda \beta - \mu \beta)^2} \]

\[ + \frac{2\lambda(\beta^2 + \lambda \alpha + \alpha \beta)(\lambda F_1(1) + \theta F_4(1) + (\theta + \phi)F_4'(1))}{2(\alpha \lambda + \lambda \beta - \mu \beta)^2} \]  \hspace{1cm} (24)

Differentiating equation (13) and taking a limit, we have

\[ F_5'(1) = \frac{a(\lambda + \beta)F_2(1) + \alpha \beta F_4'(1)}{\beta^2} \]  \hspace{1cm} (25)
The values of all probability generating functions and their derivatives are implicitly in terms of single probability $p_{0.3}$, which can be obtained from below mentioned normalization condition.

$$\sum_{n=1}^{\infty} p_{n1} + \sum_{n=0}^{\infty} p_{n2} + p_{03} + \sum_{n=0}^{\infty} p_{n4} + \sum_{n=1}^{\infty} p_{n5} = 1$$

$$F_{1}(1) + F_{2}(1) + p_{03} + F_{4}(1) + F_{5}(1) = 1$$  \hspace{1cm} (26)

4. Performance Measures

Expected orbit length $O[L] = \sum_{i=1}^{5} \text{Expected orbit length in } i\text{th state } O[L_{i}]$

$$= F_{1}'(1) + F_{2}'(1) + F_{4}'(1) + F_{5}'(1)$$

Probability of server in the busy state $Pr_{b} = F_{2}(1) + F_{4}(1)$

Probability of server in the free state $Pr_{f} = F_{1}(1) + p_{03} + F_{5}(1)$

$$= 1 - Pr_{b}$$

Availability of server A = $F_{1}(1) + F_{2}(1) + p_{03} + F_{4}(1)$

Probability of server break down $Pr_{d} = F_{5}(1)$

5. Graphical Illustrations

This section illustrates the effect of various parameters on system size and different probabilities of system states and server availability. We have also optimized the cost with respect to the slow service rate.

In the below graphs, we have fixed the parameters as $\lambda=1$, $\mu=5$, $\phi=3$, $\theta=0.3$, $\alpha=0.2$, $\beta=0.1$, $\gamma=0.7$, $\zeta=3$ unless they are used as a variable in the graph

5.1. Sensitivity analysis

Fig 1. Effect of service rate $\mu$ on Expected orbit length $O[L]$ for different values of $\alpha$. 

![Graph showing Expected Orbit Length vs. $\mu$ for different $\alpha$ values](image-url)
From Fig. 1, we observe that the expected orbit length decreases with an increasing value of $\mu$. As $\mu$ increases, the mean service time decreases, resulting in decreased mean orbit length $O[L]$. Further, we see that with a decrease in the value of $\alpha$, $O[L]$ reduces. Because with a reduced number of breakdowns in the server, Orbit length is decreased.

![Graph showing expected orbit length $O[L]$ versus retrial rate $\xi$ for different values of $\alpha$.](image)

Fig. 2. Expected orbit length $O[L]$ versus retrial rate $\xi$ for different values of $\alpha$.

Fig. 2 depicts variation in expected orbit length with change in $\xi$ for three sets of values of $\alpha$. As $\xi$ increases, the mean retrial time decreases, resulting in fast retrials, thereby decreasing the length of the orbit. This decrease becomes more obvious with a decrease in $\alpha$ as expected. The reason behind this is meantime to breakdown increases with a decrease in $\alpha$, which results in a corresponding decrease in mean orbit length.

![Graph showing probabilities of server states $Pr_b$, $Pr_f$ versus service rate $\mu$.](image)

Fig. 3. Probabilities of server states $Pr_b$, $Pr_f$ versus service rate $\mu$. 
Fig. 4. Probabilities of server states $P_{rb}$, $P_{rf}$ versus arrival rate $\lambda$.

Fig. 5. Effect of breakdown parameter $\alpha$ on server state probabilities $P_{rb}$, $P_{rf}$.

Fig. 6. Effect of $\alpha$ on availability of server $A_r$ and probability of server in a breakdown state $P_{rd}$. 
Figs. 3-5 reveal how probabilities of busy and free servers change with $\mu$, $\lambda$, and $\alpha$, respectively. As clear from Fig., with an increase in $\mu$, the mean service time decreases. Hence, the probability of the server being busy decreases; consequently, that of being free increases. In the same way, with an increase in $\lambda$, the inter-arrival time decreases, and hence the probability of busy server increases. As $\alpha$ increases, due to the corresponding decrease in the meantime to breakdown, the probability of the server remaining in busy state increases, which is verified in Fig. 5.

![Graph showing effect of $\beta$ on availability of server $A_v$ and probability of server in a breakdown state $Pr_d$.](image)

Figs. 6 and 7 depict that as $\alpha$, $\beta$ increase, the availability of servers decreases and increases respectively. This is due to a decrease in the meantime to breakdown and means repair time, respectively. Hence the complementary probability of the server being in breakdown state increases and decreases respectively with an increase of $\alpha$, $\beta$.

### 5.2. Cost analysis

In this subsection, we minimize the operating cost function with respect to $\phi$. In order to obtain the optimal value of $\phi$, we define some cost elements below:

- $C_L =$ cost per unit time for each customer present in orbit.
- $C_\mu =$ cost per unit time for service in normal state.
- $C_\phi =$ cost per unit time for service in working vacation state.
- $C_\alpha =$ cost per unit time in the vacation period.
- $C_\beta =$ cost per unit time in the breakdown state
- $C_\phi =$ cost per unit time for repair

The corresponding cost function per unit time is defined as

$$F(\phi) = O[L]C_L + \mu C_\mu + \phi C_\phi + \alpha C_\alpha + \beta C_\beta$$

To find the optimal cost $F(x)$ and the corresponding value of $x$, we take $C_L = 24$, $C_\mu = 35$, $C_\alpha = 12, C_\phi = 20, C_\alpha = 6, C_\beta = 10$ in the parabolic method. This method generates quadratic function through calculated points in every iteration to which $F(x)$ can be
approximated. The point at which \( F(x) \) is optimum in three-point pattern \( \{x_1, x_2, x_3\} \) is given by

\[
x_L = \frac{0.5(F(x_1)(x_2^2 - x_3^2) + F(x_2)(x_3^2 - x_1^2) + F(x_3)(x_1^2 - x_2^2))}{F(x_1)(x_2 - x_3) + F(x_2)(x_3 - x_1) + F(x_3)(x_1 - x_2)}
\]

The new value obtained replaces one of the three points to improve the current 3 points patterns. This process is used iteratively till the optimum value is obtained up to the desired degree of accuracy.

Table 1 shows that optimum value \( F(\phi) = 393.054518 \) correspondings to \( \phi = 3.785980 \) with the permissible error of \( 10^{-5} \), which agrees with the results of Fig. 8.

Table 1. The parabolic method to optimize the cost function.

<table>
<thead>
<tr>
<th>#</th>
<th>( \phi_1 )</th>
<th>( \phi_2 )</th>
<th>( \phi_3 )</th>
<th>( f(\phi_1) )</th>
<th>( f(\phi_2) )</th>
<th>( f(\phi_3) )</th>
<th>( \phi_L )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.00</td>
<td>4.00</td>
<td>5.00</td>
<td>395.157397</td>
<td>393.187097</td>
<td>396.769435</td>
<td>3.85484</td>
</tr>
<tr>
<td>2</td>
<td>3.00</td>
<td>3.85484</td>
<td>4.00</td>
<td>395.157397</td>
<td>393.068554</td>
<td>393.187097</td>
<td>3.80217</td>
</tr>
<tr>
<td>3</td>
<td>3.00</td>
<td>3.802176</td>
<td>3.85484</td>
<td>395.157397</td>
<td>393.055302</td>
<td>393.068554</td>
<td>3.79106</td>
</tr>
<tr>
<td>4</td>
<td>3.00</td>
<td>3.791063</td>
<td>3.802176</td>
<td>395.157397</td>
<td>393.054596</td>
<td>393.055302</td>
<td>3.78724</td>
</tr>
<tr>
<td>5</td>
<td>3.00</td>
<td>3.787246</td>
<td>3.791063</td>
<td>395.157397</td>
<td>393.054522</td>
<td>393.054596</td>
<td>3.78634</td>
</tr>
<tr>
<td>6</td>
<td>3.00</td>
<td>3.78634</td>
<td>3.787246</td>
<td>395.157397</td>
<td>393.054518</td>
<td>393.054522</td>
<td>3.78605</td>
</tr>
<tr>
<td>7</td>
<td>3.00</td>
<td>3.786053</td>
<td>3.78634</td>
<td>395.157397</td>
<td>393.054517</td>
<td>393.054518</td>
<td>3.78598</td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper, \( M/M/1 \) queuing system with multiple working vacations, vacation interruption, waiting server, breakdown, and repair under classical retrial policy is studied. Using the probability generating function technique, we have obtained the closed-form expressions of mean system sizes and probabilities of different server states. The orbit size is found to decrease with an increase in service and retrial rate. It is found to further decrease with an increase in breakdown parameters. The variation in probabilities of
different server states with arrival rate, service rate, and breakdown parameter is investigated. The availability of servers with variation in breakdown and repair rate is also analyzed. The graphical behavior is found in accordance with the theoretical expectations. The optimal value of slow service rate for the expected cost of the model is obtained via the quadratic fit approach, and the obtained optimal slow service rate is found to validate the results obtained graphically. For future scope, a similar model with batch arrival can be studied.

References