Transient Analysis of a Finite Capacity M/M/1 Queuing System with Working Breakdowns and Recovery Policies

Nneamaka Judith Ezeagu
Department of Mathematics, Pan African University, Institute for Basic Sciences, Technology and Innovations (PAUISTI), Nairobi, 62000-00200, Kenya.

George Otieno Orwa
School of Mathematical Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, 62000-00200, Kenya.

Michael Johannes Winckler
Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Heidelberg, 69120, Germany.

Abstract
In this study, a finite capacity M/M/1 queuing system with working breakdowns and recovery policies is modeled by a set of differential equations. The recovery policies are divided into two based on the recovery periods: $k$-threshold and the startup recovery. The $k$-threshold recovery is assumed to start after a certain number $k \geq 1$ of customers arrive in the system and the the start-up recovery can only be done when the system is empty. It is shown that the model satisfies the linear conservation law and therefore preserves the total law of probability. By employing the fourth order Runge-Kutta method, the time-dependent state probabilities were obtained. Further, the system performance measures were analyzed to understand the sensitivity of the model parameters. Numerical results show that by using a smaller $k$-threshold value, adjusting the service capacity and repair rates accordingly to balance respectively the service demand and breakdowns greatly improves the system performance.

AMS subject classification: M/M/1, Queuing system, Working breakdowns, Recovery policies, Quasi birth-and-death process, Time-dependent probabilities, Numerical solution.

Keywords: M/M/1, Queuing system, Working breakdowns, Recovery policies, Quasi birth-and-death process, Time-dependent probabilities, Numerical solution.

1Corresponding author: E-mail: nneamaka23@gmail.com
1. Introduction

A queueing system is one in which customers arrive to access a facility for a given service, wait if the facility is busy or unavailable and leave after being served. Some practical examples of a queueing system include the telecommunication-network and the road systems. These systems usually operate under unpredictable (random) environmental conditions, and queues are formed due to the unbalanced services and random service demand. The random environmental conditions are mainly classified according to the behavior (state) of the service mechanism (server) at any point in time. The server states include: server in a normal working state, vacation, working vacation (WV), breakdowns, or working breakdown states. Different queueing models have been used by several authors to model and analyze queueing systems under the aforementioned server states. This paper focuses its review on the assumptions of a Poisson arrival, Markovian exponential service times and single server (M/M/1) queuing model.

A server according to [1, 2] is on vacation when the system becomes unavailable for services for a period of time. Kalidass and Ramanath [3], Ibe and Isijola [2] studied an M/M/1 queue with server vacation. The authors analyzed the system under the steady state condition and the former, also studied the time-dependent solution. When a server is interrupted on vacation or is set to provide service during the vacation state at a slower rate then the server is said to be on a WV (see [4]). The M/M/1 queue with WV was first introduced by Servi and Finn [4]. Subsequently, Li and Tian [5] incorporated the vacation interruption policy, and later, Isijola and Ibe [1] extended the model to include two types of vacation interruptions: partial interruption and complete interruption policy. The latter authors assumed that if the number of customers in the system reaches the threshold set for each vacation type, then the vacation can be interrupted. Yang et al. [6] studied the F-policy M/M/1/K queuing system with WV and an exponential start-up time. The F-policy is assumed to control arrivals in the sense that when arrivals reaches the system capacity, customers are not allowed to enter the system until the queue length decreases to a certain threshold value. In some queueing systems, servers randomly breakdown while customers are on service. Due to server breakdowns, customers’ services are completely stopped until the server is repaired or replaced by another server [7, 8]. Ke and Pearn [8] studied a management policy on an M/M/1 queue for heterogeneous arrival with server breakdowns and vacation. Server vacation was considered as a removable server which operates on an N policy. This according to the study is such that when \( N \geq 1 \) or more units are in the system, the server is turned on, otherwise, the server is turned off and goes on vacation. It was assumed that the server break down follow a Poisson process and the repair time has an exponential distribution. Gray et al. [7] analysed a multiple-vacation queuing model subject to server breakdowns while in operation. An N-policy M/M/1 queuing system with WV and server breakdowns was investigated by Yang and Wu [9].

Kalidass and Kasturi [10] introduced the M/M/1 queuing system with random working breakdowns and repairs while in operation. Service during a working breakdown state is not completely stopped but rather goes at a slower rate compared to that in a normal state. Liu and Song [11] extended the work in [10] to batch arrival queue, and
Jiang and Xin [12] modified the idea to consider repair as delaying under a Bernoulli-schedule-controlled policy. Jiang and Xin classified the service process as regular and auxiliary service period. As assumed by the authors, at breakdown during the regular period, the system may be subjected to a repair with probability $p$ or continue to provide auxiliary service with probability $1 - p$. Once a breakdown occurs in the auxiliary period the system immediately goes through repair with probability one. By the use of matrix analytic method, the steady state distribution was computed. Yang [13] modified the work in [10] by incorporating server vacations, and employed the Runge-Kutta fourth (RK4) order scheme to obtain the time-dependent state probabilities. Similar numerical scheme was used in [14, 15] to obtain the transient state probabilities and finally compute the performance measure. Recently, Yang et al. [16] also used the RK4 scheme to obtain the time-dependent state probabilities of a finite-capacity M/M/1 queuing system with working breakdowns, reneging, and retention of impatient customers. The corresponding steady state probabilities were computed using the matrix analytic approach. It is important to any system manager to ensure that an acceptable level of service is provided in terms of response time requirements while avoiding excessive cost. In regards to this, threshold-based recovery has found its application in many areas of queuing system. According to the recovery policy as first introduced in Efrosinin and Semenova [17], repair can only be considered for a break down server if the number of arriving customers exceeds a certain threshold. The authors studied an infinite M/M/1 queuing system with an unreliable device. The device is said to completely fail and recover in an exponentially distributed time. The authors used the probability generating functions (PGFs) to analyze the stationary performance of the system. With similar assumptions, Yang et al. [18] analyzed a finite capacity M/M/1 system with a threshold-based recovery policy under the steady state condition.

Based on the reviewed literatures, server repair is assumed to start immediately the server breaks down. However, it is not always true for situations where service revenue for certain number of arriving customers is less than the repair cost. Perhaps, in many practical applications, repair can only be considered for a working breakdown server when the system is empty or a considerably increase of the arriving customers is observed. However, not much has been done on M/M/1 queuing system with working breakdowns. This therefore inspires us to investigate a single server finite capacity queuing system with working breakdowns and recovery policies motivated by the study in [10, 17] and [18]. In addition to that, in queuing systems where the service time is limited, the system never approaches equilibrium state. The need for time-dependent analysis for such a system is important to understand the transient system behavior. This is also more important to managers in many queuing applications where the knowledge of the time-dependent behavior of the system is necessary rather than the steady-state results. This study thus focuses its analysis on the time-dependent performances by means of a numerical approach.
2. Methods

2.1. Poisson Distribution

The Poisson distribution involves the probability of occurrence of events that can be counted as integer numbers. It is defined as

\[ P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, 2, \ldots, \]  

where \( \lambda \) is the rate parameter, \( n \) is the number of events observed in a given time interval \( t \geq 0 \). In this study, we consider only one event can occur between the time \( t \) and \( t + h \), where \( h \) is an infinitesimal time. Therefore,

\[ P_0(h) = e^{-\lambda h} \approx 1 - \lambda h + o(h), \]  

\[ P_1(h) = \lambda he^{-\lambda h} \approx \lambda h + o(h), \]  

\[ P_{n\geq2}(h) = o(h^2), \]  

where \( o(h^2) \ll o(h) \) with

\[ \lim_{h \to 0} \frac{o(h)}{h} = 0. \]  

2.2. Exponential Distribution

Let \( T \) be a random variable that counts the time an event is completed. Given that \( \alpha \) is the rate parameter, then by the Markov (i.e., memoryless) property of exponential distribution, we have,

\[ P[T > t + h] = P[T > h] = e^{\alpha h}, \quad h > 0. \]  

In other words, if no event has been produced in \( t \) units of time, then the distribution of an additional time \( h \) is the same as it would be if no event time has passed. This simply means the system does not remember that \( t \) time units produced no event. Thus by the total law of probability, the probability that an event occurs before time \( h \) is given by

\[ P[T \leq h] = 1 - e^{\alpha h} \approx \alpha h + o(h) \]  

2.3. Quasi Birth-and-death Process

The formulation of the model proposed in this study is based on the Quasi birth-and-death (QBD) process. A QBD process is a Markov chain defined on the state space, say \( \{ (i, j) : 0 \leq i \leq n_j, j \in \mathbb{N} \cup \{0\} \} \) where the state space is divided into levels such that for each level \( j \), there are \( n_j \) states. This process allows transitions only to the neighboring levels or within the same level as the queuing system describes. In this study, birth is apportioned to an event counted “into” the present state and death as an event counted “out” of the present state.
2.4. Model Formulation

Consider an M/M/1 queuing system (for computation purpose, we assume a finite capacity system of \( N < \infty \) levels) with working breakdowns and recovery policies while in operation. We divide the recovery policies into two according to the recovery periods: \( k \)-threshold recovery and the startup recovery. The \( k \)-threshold recovery assumes that server repair starts after a certain number \( k \geq 1 \) customers arrive in the system and the startup recovery is defined as a recovery carried out only when the system is empty. Initially, the system starts operation with a working normal server, and customers arrive independently following a Poisson distribution at a constant arrival rate \( \lambda \). Server breakdowns occur only when there is at least one customer in the system and follows a Poisson process at a rate \( \alpha \). Therefore, upon arrival of a customer, a normal working server can either provide service at an exponentially distributed service time with rate \( \mu_w \) or breaks down. Broken down server is allowed to provide service at an exponentially distributed service times at a rate \( \mu_b \) such that \( \mu_b < \mu_w \) or goes through the \( k \)-threshold recovery or the startup recovery. The \( k \)-threshold recovery and the startup recovery are exponentially distributed respectively with rates \( \beta \) and \( \beta_0 \). Let \( S(t) \) denote the state of the server at any instant of time such that

\[
S(t) = \begin{cases} 
1, & \text{if the server is on working normal state,} \\
0, & \text{if the server is on working breakdown state.}
\end{cases}
\]

Therefore, the server follows a Markovian environmental process \( \{S(t), t \geq 0\} \) on a finite state space \( \{1, 0\} \). Let \( N(t) \) denote the total number of customers in the system at any time \( t \). Thus, the system describes a continuous time Markov chain \( \{S(t), N(t), t \geq 0\} \) on the state space \( \{(i, n): i \in \{1, 0\}, n = \{0, 1, \ldots, N\}\} \). We define the time-dependent state probability

\[
P_{i,n}(t) = P[S(t) = i, N(t) = n]; \quad \forall \ i = 1, 0; n = \{0, 1, \ldots, N\}
\]

(7)

as the probability that for each server state \( i \), there are \( n \) customers in the system at any time, \( t \).

Remark

In the rest of this paper we would use the letter subscripts “w” and “b” to denote respectively the “working normal” server and the “working break down” server such that \( P_{w,n}(t) = P_{1,n}(t) \) and \( P_{b,n}(t) = P_{0,n}(t) \) and so on, unless otherwise stated.

Hence, using the QBD process (see section 2.3) and substituting the probabilities given in (2)-(4), and (6), and applying the definition given in (5), we obtain the following differential equations governing the M/M/1 queuing system with working breakdowns
and recovery policies as:

\[
\frac{dP_{w,0}}{dt} = -\lambda P_{w,0}(t) + \mu w P_{w,1}(t) + \beta_0 P_{b,0}(t), \quad (8)
\]

\[
\frac{dP_{w,n}}{dt} = \lambda P_{w,n-1}(t) - (\lambda + \mu w + \alpha) P_{w,n}(t) + \mu w P_{w,n+1}(t); \quad n = 1, 2, \ldots, k, \quad (9)
\]

\[
\frac{dP_{w,n}}{dt} = \lambda P_{w,n-1}(t) - (\lambda + \mu w + \alpha) P_{w,n}(t) + \mu w P_{w,n+1}(t) + \beta P_{b,n}(t);
\]

\[n = k + 1, \ldots, N - 1, \quad (10)\]

\[
\frac{dP_{w,N}}{dt} = \lambda P_{w,N-1}(t) - (\mu w + \alpha) P_{w,N}(t) + \beta P_{b,N}(t), \quad (11)
\]

\[
\frac{dP_{b,0}}{dt} = -(\lambda + \beta_0) P_{b,0}(t) + \mu b P_{b,1}(t), \quad (12)
\]

\[
\frac{dP_{b,n}}{dt} = \lambda P_{b,n-1}(t) - (\lambda + \mu b) P_{b,n}(t) + \mu b P_{b,n+1}(t) + \alpha P_{w,n}; \quad n = 1, 2, \ldots, k, \quad (13)
\]

\[
\frac{dP_{b,n}}{dt} = \lambda P_{b,n-1}(t) - (\lambda + \mu b + \beta) P_{b,n}(t) + \mu b P_{b,n+1}(t) + \alpha P_{w,n};
\]

\[n = k + 1, \ldots, N - 1, \quad (14)\]

\[
\frac{dP_{b,N}}{dt} = \lambda P_{b,N-1}(t) - (\mu b + \beta) P_{b,N}(t) + \alpha P_{w,N}(t), \quad (15)
\]

with the initial state probabilities given by

\[P_{w,0}(0) = 1; \quad P_{w,n}(0) = 0, \quad \forall \ n = 1, 2, \ldots N; \quad P_{b,n}(0) = 0, \quad \forall \ n = 0, 1, 2, \ldots, N. \quad (16)\]

3. Model Properties and Performance Measures

3.1. Model Basic Properties

3.1.1 Transition Rate Matrix

This is a square matrix of real numbers describing the rate at which the chain of the queuing process moves from one state to the other. Let \(Q = (q_{i,j} : i, j \in I)\) be the transition matrix, where \(I\) is a countable index set. Then \(Q\) is derived such that the diagonal entries, \(q_{i,i}\), satisfy the row sum \(q_{i,i} = -\sum_{j \neq i}^m q_{i,j} < \infty\) where \(q_{i,j}, i \neq j\) are
the off-diagonal entries \((m = 2N + 2)\) state variables. Thus

\[
Q = \begin{bmatrix}
0 & \begin{bmatrix} A_0 & \Lambda \\
M & A_1 & \Lambda \\
& \ddots & \ddots & \ddots \\
& & M & A_2 & \Lambda \\
& & & \ddots & \ddots & \ddots \\
& & & & & M & A_3 \end{bmatrix} \\
1 & \begin{bmatrix} 0 & 0 \\
M & A_1 & \Lambda \\
& \ddots & \ddots & \ddots \\
& & M & A_2 & \Lambda \\
& & & \ddots & \ddots & \ddots \\
& & & & & M & A_3 \end{bmatrix} \\
\vdots & \vdots \\
\begin{bmatrix} k & \Lambda \\
M & A_1 & \Lambda \\
& \ddots & \ddots & \ddots \\
& & M & A_2 & \Lambda \\
& & & \ddots & \ddots & \ddots \\
& & & & & M & A_3 \end{bmatrix} \\
\begin{bmatrix} k+1 & \Lambda \\
M & A_1 & \Lambda \\
& \ddots & \ddots & \ddots \\
& & M & A_2 & \Lambda \\
& & & \ddots & \ddots & \ddots \\
& & & & & M & A_3 \end{bmatrix} \\
\vdots & \vdots \\
\begin{bmatrix} N-1 & \Lambda \\
M & A_1 & \Lambda \\
& \ddots & \ddots & \ddots \\
& & M & A_2 & \Lambda \\
& & & \ddots & \ddots & \ddots \\
& & & & & M & A_3 \end{bmatrix} \\
N & \begin{bmatrix} N & \Lambda \\
M & A_1 & \Lambda \\
& \ddots & \ddots & \ddots \\
& & M & A_2 & \Lambda \\
& & & \ddots & \ddots & \ddots \\
& & & & & M & A_3 \end{bmatrix}
\end{bmatrix}
\]

Figure 1: Transition rate matrix

where

\[
A_0 = S_0 - \Lambda, \quad A_1 = S_1 - M - \Lambda, \quad A_2 = S_2 - M - \Lambda, \quad \text{and} \quad A_3 = S_2 - M.
\]

The matrix \(Q\) is a \((2N + 2) \times (2N + 2)\) square matrix whereby the set of state levels \(\{(i, j), i \in \{1, 0\}, 0 \leq j \leq N\}\) is represented by \(j\). The entries \((S_i)_{j=0,1,2}\), \(M\) and \(\Lambda\) are again matrices of order 2 (i.e., the number of environmental states). The sub-matrices \(S_i\) are the infinitesimal generator of the environmental process \(\{S(t), t \in [0, \infty)\}\) given with values as follows:

\[
S_0 = \begin{bmatrix} 0 & 0 \\ \beta_0 & -\beta_0 \end{bmatrix}, \quad S_1 = \begin{bmatrix} -\alpha & \alpha \\ 0 & 0 \end{bmatrix}, \quad S_2 = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix},
\]

while \(M\) and \(\Lambda\) are diagonal matrices listed below

\[
M = \begin{bmatrix} \mu_w & 0 \\ 0 & \mu_b \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}.
\]

Observe that the matrix \(Q\) is of a block-tridiagonal structure and have the characteristic of the Markov processes described in [19].

### 3.1.2 Linear Conservation Law

We can see that the model differential equations (8)–(15) is linear in the state dependent variables. Therefore, we let \(P\) be a real-valued column vector containing the unknown time-dependent state probabilities. Then, the model can be written in a general linear form

\[
\begin{cases}
P'(t) = P(t)Q, \\
P(0) = A.
\end{cases}
\]

(17)
where $Q$ is the transition rate matrix, $F$, the vector containing the known functions (given by the right hand side of the model differential (8)–(15)), and $A = [1 \ 0 \ \cdots \ 0]^T$ is a vector of the initial values (16). Given any constant column vector $V = [v_i,1]_{(2N+2) \times 1}$ and in particular, $v_i,1 = 1, \forall i$, we have that

$$V^T F(P(t)) \equiv 0,$$

$$\implies V^T P'(t) = V^T F(P(t)) = 0$$

Thus, the model (17) satisfies the linear conservation law

$$V^T P(t) = V^T A. \quad (18)$$

Hence, it follows that

$$V^T P(t) \equiv 1$$

which implies that

$$\sum_{j=0}^{N} P_{w,j}(t) + \sum_{j=0}^{N} P_{b,j}(t) = 1. \quad (19)$$

Equation (19) is the normalization condition of the state time-dependent probabilities of the model (17). This therefore shows that the law of probability is preserved by the linear conservation law (18).

4. Model Performance Measures

Once the time-dependent probabilities of the system are obtained, we evaluate the system behaviour based on the following performance measures.

4.1. System State Probabilities

The system state probabilities will be evaluated to understand how the probabilities of the number of customers in the system are distributed. Thus, the total system probability defined as the probability of having $n$ customers in the system at any time $t \geq 0$ is given by

$$P_n(t) = P_{w,n}(t) + P_{b,n}(t). \quad (20)$$

with $P_{w,n}(t)$ and $P_{b,n}(t)$ respectively the time-dependent working normal and working breakdown state probabilities.
4.2. Expected Number of Customers in the System

This is defined as the average number of customers waiting in queue prior to service written mathematically as

\[ \mathbb{E}[N(t)] = \mathbb{E}[N_w(t)] + \mathbb{E}[N_b(t)], \]  

(21)

where

\[ \mathbb{E}[N_w(t)] = \sum_{n=0}^{N} n P_{w,n}(t), \]  

(22)

is the expected number of customers in the working normal state and

\[ \mathbb{E}[N_b(t)] = \sum_{n=0}^{N} n P_{b,n}(t), \]  

(23)

is the expected number of customers in the working breakdown state.

**Remark**

In the following section, we write \( P_n, P_{w,n}, P_{b,n}, \mathbb{E}[N], \mathbb{E}[N_w] \) and \( \mathbb{E}[N_b] \) to read \( P_n(t), P_{w,n}(t), P_{b,n}(t), \mathbb{E}[N(t)], \mathbb{E}[N_w(t)] \) and \( \mathbb{E}[N_b(t)] \) respectively, which are to be used only in the graphs and legends.

5. Results and Discussion

It is the focus of this paper to analyze the transient behavior of a finite capacity M/M/1 queuing system modeled by the differential (8)-(15). In regards to this, we employ the fourth order Runge-Kutta numerical scheme to obtain the time-dependent state probabilities and further compute the performance measures. The numerical scheme was chosen since it preserves the linear conservation law and consequently preserves the model property. This section therefore presents the numerical results summarized with different plots under varying parameter values. For ease of computation, we fix the system capacity \( N = 15 \), time to range from \( t = 0 \) to \( t = 120 \), and the standard parameter values: \( \lambda = 0.35, \mu_w = 0.5, \mu_b = 0.2, \alpha = 0.15, \beta_0 = 0.05, \beta = 0.2 \). Figure 2 shows the time-dependent total system size probability distribution, \( P_n(t) \), of the number of customers in the system. A few probability curves are displayed to clearly understand the distribution trend of the system probabilities over the specified time interval.

It is seen that a few number of customers are most likely to be seen at the early time the system starts operation. Whereas, in the long run as the system evolves, the less likely it is to have a few customers while the chances of having a full capacity system increases. Having seen from fig. 2, how the probability is distributed as the system evolves over time, we therefore, deem it important to analyze what is expected of the
system probabilities at $t = 120$. In regards to that, the histograms shown in figure 3 are plotted to depict the probability distribution of the number of customers at $t = 120$. Each histogram displayed in figures 3a–3d shows that, before the start of the $k$-threshold recovery, the probability increases with an increase in the number of customers $n$ until the $k$-threshold value is reached. We see that immediately after the start of the threshold recovery, the probability decreases with an increase in the number of customers, $n$. This trend continues until $N - 1$ customers are in the system, then, a slight increase occurs in the probability of having $N = 15$ customers. This can be explained to be that at this state, the system capacity is reached and hence there is no transition (arrival or service) to (or from) the next state. It is observed for each $k$-threshold value that the most likely occurrences of the number of customers in the system are clustered just around the $k$ value, with the highest probability lying on the $k$ value. On comparing the performances of the $k$-threshold recovery values in figures 3a–3d, we see that for $k = 3, 6, 9$ and 12, the probability of having an empty system is approximately 8.1%, 3.1%, 1.2%, and 0.6% respectively. While on the other hand, the probability of having a full system of 15 customers is approximately 3.5%, 5.1%, 7.6%, and 11.8% respectively. In other words, the chances of getting the customers served on time increases for a smaller $k$-threshold recovery value.
Figure 3: Effect of varying the $k$-threshold recovery value on the probability distribution of the number of customers in the system at $t = 120$. Each histogram 3a, 3b, 3c and 3d corresponds to a specific value of $k = 3, 6, 9$ and 12 respectively. All histograms are produced for all values of $n$, and all other standard parameter values remained constant.

In figure 4, it is observed that for each plot, $E[N(t)]$ increases as time evolves. Starting from $t = 0$ until $t = 10$, all plots lie on the same points and thus indicates that $E[N(t)]$ is approximately the same for all $k = 3, 6, 9$ and 12 at that time interval. As each plot corresponding to a particular value of $k$ diverges out, we see that $E[N(t)]$ reduces with a smaller threshold value, and increases with a large threshold value. This indicates that the earlier the threshold recovery is initiated, the higher the chances of having all customers served on time. Therefore, having observed that the system performs better with a small value of $k$, we subsequently analyze the system for the 3-threshold recovery value. In figure 5, we see that increase in $\beta$ increases the probability of having no customers in the system and also decreases the probability of having 15 customers at that instant of time.

Tables 1–3 present numerical results for easy comparison between the working normal and working breakdown probability distribution. It is clear from Table 1 that increase in the recovery rate $\beta$, increases $P_{w,0}$, and $P_{b,0}$, and at the same time decreases $P_{w,15}$, and $P_{b,15}$. In the same way, we see that an increase in the breakdown rate $\alpha$ reduces
Figure 4: Effect of the threshold recovery value, $k$ on the time-dependent expected number of customers, $E[N(t)]$ in the system. Each plot corresponds to a specific value of $k = 3, 6, 9$ and 12 while the other standard parameter values remained fixed.

Figure 5: Effect of the repair rate, $\beta$ on the probability distribution of the number of customers in the system at $t = 120$. Each plot was produced for the 3-threshold recovery system, and for all values of $n$ with the breakdown rate $\alpha = 0.17$ while the other standard parameter remained constant.
Table 1: Effect of the breakdown rate $\alpha$ and the repair rate $\beta$ on the probability distribution of $n = 0$ and $n = 15$ customers in the system at $t = 120$ ($k = 3$ and other standard parameters are constant).

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$P_{w,0}$</th>
<th>$P_{b,0}$</th>
<th>$P_0$</th>
<th>$P_{w,n}$</th>
<th>$P_{b,n}$</th>
<th>$P_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = 0.10$</td>
<td>0.0588</td>
<td>0.0168</td>
<td>0.0756</td>
<td>0.0319</td>
<td>0.0416</td>
<td>0.0465</td>
</tr>
<tr>
<td>$\beta = 0.15$</td>
<td>0.0181</td>
<td>0.0385</td>
<td>0.0567</td>
<td>0.0228</td>
<td>0.0641</td>
<td>0.0869</td>
</tr>
<tr>
<td>$\beta = 0.20$</td>
<td>0.0139</td>
<td>0.0184</td>
<td>0.0323</td>
<td>0.0191</td>
<td>0.0340</td>
<td>0.0530</td>
</tr>
<tr>
<td>$\beta = 0.25$</td>
<td>0.0109</td>
<td>0.0101</td>
<td>0.0210</td>
<td>0.0156</td>
<td>0.0199</td>
<td>0.0355</td>
</tr>
</tbody>
</table>

$P_{w,0}$ and $P_{b,0}$, and obviously increases $P_{w,15}$, and $P_{b,15}$. On a closer observation, we can see that when $\alpha$ is much higher than $\beta$, the system is more often in the breakdown state and therefore, $P_{b,15}$ gets higher than for $P_{b,0}$. Table 2 simply shows that $P_{w,0}$, $P_{b,0}$ and $P_0$ increase with an increase in $\mu_w$ and $\beta$. In Table 3, it is clear that an increase in $\lambda$ increases $P_{w,15}$, $P_{b,15}$ and $P_{15}$ and reduces $P_{w,0}$, $P_{b,0}$ and $P_0$. Thus, however, in Tables 1–3, $P_{w,0}$ is much higher than $P_{b,0}$ whereas $P_{w,15}$ is less than $P_{b,15}$.

Table 2: Effect of the normal service rate $\mu_w$ and the repair rate $\beta$ on the probability distribution of $n = 0$ and $n = 15$ customers in the system at $t = 120$ ($\alpha = 0.17; k = 3$), and other standard parameters are constant.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\mu_w = 0.45$</th>
<th>$\mu_w = 0.5$</th>
<th>$\mu_w = 0.52$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{w,0}$</td>
<td>$P_{b,0}$</td>
<td>$P_0$</td>
<td>$P_{w,n}$</td>
</tr>
<tr>
<td>$\beta = 0.10$</td>
<td>0.0164</td>
<td>0.0104</td>
<td>0.0269</td>
</tr>
<tr>
<td>$\beta = 0.15$</td>
<td>0.0257</td>
<td>0.0136</td>
<td>0.0393</td>
</tr>
<tr>
<td>$\beta = 0.20$</td>
<td>0.0241</td>
<td>0.0740</td>
<td>0.0981</td>
</tr>
<tr>
<td>$\beta = 0.25$</td>
<td>0.0214</td>
<td>0.0681</td>
<td>0.0895</td>
</tr>
</tbody>
</table>

Table 3: Effect of the arrival rate $\lambda$ and the normal service rate $\mu_w$ on the probability distribution of $n = 0$ and $n = 15$ customers in the system at $t = 120$ ($\alpha = 0.17; k = 3$), and other standard parameters remained constant.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\mu_w = 0.45$</th>
<th>$\mu_w = 0.5$</th>
<th>$\mu_w = 0.52$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_w$</td>
<td>$P_{w,0}$</td>
<td>$P_{b,0}$</td>
<td>$P_0$</td>
</tr>
<tr>
<td>$\mu_w = 0.45$</td>
<td>0.0649</td>
<td>0.0383</td>
<td>0.1032</td>
</tr>
<tr>
<td>$\mu_w = 0.5$</td>
<td>0.0886</td>
<td>0.0114</td>
<td>0.0201</td>
</tr>
<tr>
<td>$\mu_w = 0.52$</td>
<td>0.0986</td>
<td>0.0472</td>
<td>0.1458</td>
</tr>
<tr>
<td>$\mu_w = 0.52$</td>
<td>0.0040</td>
<td>0.0061</td>
<td>0.0101</td>
</tr>
</tbody>
</table>
Figure 6: Effect of the breakdown rate $\alpha$ and the repair rate $\beta$ on the expected value of the number of customers ($\mathbb{E}[N(t)]$) in the system at any time. Each plot is produced for the $k = 3$-threshold recovery value system. The blue lines corresponds to the low breakdown rate and the dotted lines correspond to the higher repair rates.

Figure 7: Effect of the arrival rate $\lambda$ and the normal service rate $\mu_w$ on the expected value of the number of customers $\mathbb{E}[N(t)]$ in the system at any time. The plots are produced for the $k = 3$-threshold recovery value system with the other standard parameter values. The black lines corresponds to the lower arrival rate and the broken lines correspond to the straight lines corresponds to the lower arrival rates.
Figure 6 shows that the expected number of customers, $E[N(t)]$ increases with an increasing $\alpha$ and decreases as $\alpha$ decreases. It can also be observed that an increase in $\beta$ results to a decrease in $E[N(t)]$, and a decrease in $\beta$ decreases $E[N(t)]$. Figure 7 shows the effect of $\lambda$ and $\mu_w$ on $E[N(t)]$ at any time. The figure shows that $E[N(t)]$ increases with increase in $\lambda$ and decreases with a decrease in $\lambda$. It is also observed that the expected number of customers in the system decreases as the $\mu_w$ increases and decreases otherwise.

6. Conclusions

This study investigated a finite capacity M/M/1 queuing system with working breakdowns and recovery policies. Servers are assumed to still render services when on breakdown state but at a slower rate as compared to when on a normal working condition. The recovery policies are divided into two based on the each recovery periods: the $k$-threshold recovery and the startup recovery. We assumed that the threshold recovery can only be initiated when a certain number $k \geq 1$ customers are in the system while the startup can only be carried out when the system is entirely empty. The governing differential equations of the queuing system were formulated using the birth-and-death process. It was shown that the model differential equations satisfy the linear conservation law, thereby, conserving the total law of probability. Based on the Runge-Kutta fourth order method which was implemented in MatLab, the time-dependent system size probabilities were obtained. We further analyzed the transient system performance measures to understand the sensitivity of the model parameters. The numerical results showed that:

- The system performs better with a small threshold value $k$, and thus, it would be wise to initiate an early threshold recovery.

- An increase in the recovery rate $\beta$ increases the chances of having an empty system both in the working normal and working breakdown state. This consequently, reduces the chances of having a full capacity system.

- Increase in the arrival rate $\lambda$ increases the expected number of customers, $E[N(t)]$ and increase in the normal service rate, $\mu_w$ decreases $E[N(t)]$. Therefore, improving the service mechanism such that the service rate balances with the service demand positively improves the system performance.

- Increase in the break down rate, $\alpha$ with less recovery puts the system more often in the breakdown state, thereby, increasing $E[N(t)]$. Whereas increase in the repair rate $\beta$ which must be higher that $\alpha$, recovers the system to a normal working state and leads to decrease in $E[N(t)]$. Therefore, measures need to be taken to ensure system is always or more often in the working normal state.

Acknowledgement

Nneamaka Judith specially thank the HGS MathComp, Heidelberg University for the post-bachelor learning platform provided in the course of this work, and the Pan African
University Institute for Basic Sciences, Technology and Innovation (PAUISTI) for awarding
the study scholarship.

References

[1] Isijola-Adakeja OA, Ibe OC (2014) M/m/1 multiple vacation queueing systems with

[2] Ibe OC, Isijola OA (2014) M/m/1 multiple vacation queueing systems with differ-

server vacations and a waiting server in *Proceedings of the 6th International Con-


[5] Li J, Tian N (2007) The m/m/1 queue with working vacations and vacation inter-

trolling arrivals in the queueing system with single working vacation. *Journal of


ing systems with server breakdowns and vacations. *Quality Technology & Quanti-

with n-policy and server breakdowns. *Computers & Industrial Engineering* 82:151–
158.


[11] Liu Z, Song Y (2014) The m x/m/1 queue with working breakdown. *RAIRO-

downs and delaying repair under a bernoulli-schedule-controlled policy. *Communications in Statistics-Theory and Methods* pp. 1–16.

with working breakdowns and server vacations in *Proceedings of the International
Multi-Conference of Engineers and Computer Scientists*. Vol. 2.

[14] Purohit G, Jain M, Rani S (2012) M/m/1 retrial queue with constant retrial pol-
icy, unreliable server, threshold based recovery and state dependent arrival rates.


