Multiple vacation policy for $M^X/H_k/1$ queue with un-reliable server

Madhu Jain$^1$, Richa Sharma$^{2*}$ and Gokul Chandra Sharma$^3$

Abstract

This paper studies the operating characteristics of an $M^X/H_k/1$ queueing system under multiple vacation policy. It is assumed that the server goes for vacation as soon as the system becomes empty. When he returns from a vacation and there is one or more customers waiting in the queue, he serves these customers until the system becomes empty again, otherwise goes for another vacation. The breakdown and repair times of the server are assumed to follow a negative exponential distribution. By using a generating function, we derive various performance indices. The approximate formulas for the probability distribution of the waiting time of the customers in the system by using the maximum entropy principle (MEP) are obtained. This approach is accurate enough for practical purposes and is a useful method for solving complex queueing systems. The sensitivity analysis is carried out by taking a numerical illustration.

Keywords: Batch arrival; k-type hyper-exponential distribution; State-dependent rates; Maximum entropy principle; Long-run probabilities; Un-reliable server; Queue length

Background

Server vacation models are useful for queueing systems in which the server wants to utilize his idle time for different purposes. The vacation mechanism considered in this paper is termed as ‘multiple vacation policy’. That is, the server upon returning from a vacation leaves immediately for another one if the system is empty at that moment. Applications of the server with multiple vacation models can be found in manufacturing systems, designing of computer and communication systems, etc. Queueing systems with multiple server vacations have attracted the attention of numerous researchers. Baba (1986) studied batch-arrival $M^X/G/1$ queueing systems with multiple vacations. A discrete-time Geo/G/1 queue with multiple vacations was studied by Tian and Zhang (2002). An $M^X/G(a,b)/1$ queue with multiple vacations including closedown time has been studied by Arumuganathan and Jeykumar (2004). Kumar and Madheswari (2005) analyzed a Markovian queue with two heterogeneous servers and multiple vacations. By using the matrix geometric method, they derived the stationary queue length distribution and mean system size. Wu and Takagi (2006) investigated an $M/G/1$ queue with multiple vacations and exhaustive service discipline such that the server works with different rates rather than completely stopping the service during vacation. Ke (2007) studied an $M^X/G/1$ queuing system under a variant vacation policy where the server takes at most $j$ vacations. He derived the system size distribution as well as waiting time distribution in the queue. Ke and Chang (2009) considered an $M^X/(G_1,G_2)/1$ retrial queue with general retrial times, where the server provides two phases of heterogeneous service to all customers under Bernoulli vacation schedules. They constructed the mathematical model and derived the steady-state distribution of the server state and the number of customers in the system/orbit. Ke et al. (2009) studied the vacation policy for a finite buffer $M/M/c$ queueing system with an unreliable server. Threshold $N$-policy for an $M^X/H_2/1$ queueing system with an un-reliable server and vacations was studied by Sharma (2010). Moreover, Singh et al. (2012) investigated an $M/G/1$ queueing model with vacation and used the generating function method for obtaining various performance measures. Very recently, an unreliable bulk queue with state-dependent arrival rates was examined by Singh et al. (2013). Queueing models with an un-reliable server under multiple vacation policy are more realistic representation...
of the systems. The service of the components may be interrupted when the operator encounters unpredicted breakdowns, and it is to be immediately recovered with a random time. When the repair is completed, the server immediately returns for service. Wang et al. (1999) extended Wang’s model to the N-policy for an M/H/1 queueing system and focused on single-arrival Erlangian service time queueing model with an un-reliable server. Wang et al. (2004) considered an M/H/1 queueing system with a removable and un-reliable server under N-policy and presented the optimal operating policy. Wang (2004) considered an M/G/1 queue with an un-reliable server and second optional service. Using the supplementary variable method, he obtained transient and steady-state solutions for both queueing and reliability measures of interest. Ke (2005) studied a modified T vacation policy for an M/G/1 queueing system where an un-reliable server may take at most J vacations repeatedly until at least one customer appears in the queue upon returning from vacation, and the server needs a startup time before starting each of his service periods. Li et al. (2007) proposed a single-server vacation queue with two policies, working vacation and service interruption. Choudhury and Deka (2008) studied an M/G/1 retrial queue with an additional second phase of optional service subject to breakdowns occurring randomly at any instant while serving the customers. Further, Wang and Xu (2009) obtained the solution of an M/G/1 queue with second optional service and server breakdown using the method of functional analysis. The work on an M/G/1 queue with second optional service and server breakdown has been done by Choudhury and Tadj (2009). They derived the Laplace-Stieltjes transform of busy period distribution and waiting time distribution. Further, an un-reliable server queue with multi-optional services and multi-optional vacations was analyzed by Jain et al. (2013).

Approximate results of many complex queueing models have been developed by several authors by applying the technique of maximum entropy principle. Jain and Singh (2000) used the principle of maximum entropy to analyze the optimal flow control of a G/G/c finite capacity queue. Jain and Dhakad (2003) provided the steady-state queue size distribution of an M/G/1 queue using the maximum entropy approach in which the constraints are expressed in terms of mean arrival rates, mean service rates, and mean number of customers in the system. Further, Ke and Lin (2006) employed the principle of maximum entropy to derive the approximate formulas for the steady-state probability distributions of the queue length. Ke and Lin (2008) suggested the maximum entropy principle to examine the M/G/1 queueing system in different frameworks. Omey and Gulick (2008) did maximum entropy analysis of an M/M/1 queueing model with multiple vacations and server breakdowns. Wang and Huang (2009) analyzed a single removable and un-reliable server M/G/1 queue under the (p,N)-policy. They did the maximum entropy analysis to obtain the approximate formulas for the probability distributions of the number of customers and the expected waiting time in the system. Maximum entropy approach has been applied for an un-reliable server vacation queueing model by Jain et al. (2012).

The main objective of our study is to develop an M\(^{X}/H_{\text{v}}/1\) queueing model with an un-reliable server under multiple vacation policy. In this paper, an M\(^{X}/H_{\text{v}}/1\) queue has been analyzed including more features, namely (1) bulk arrival, (2) server breakdown, and (3) multiple vacation. The various approximate results for waiting time distribution have been analyzed using the maximum entropy principle which was not considered in the previous study. Now, we cite a real-life situation of a given model wherein all the features are encountered simultaneously. To highlight the application, we cite an example of production of heat transfer equipment. In the production of these equipment, the raw material of these equipment arrives in group (batch arrival). The production of these equipment has been done by the machine in phases (called k-type hyper-exponential distribution). During the production, the production of the equipment may be interrupted due to some machinery faults called server breakdown. After interruption in the service, the machine is immediately sent for repairing. After repairing, the machine renews and works as a new one. As the raw material of these equipment is finished, the operator may take multiple vacations till the raw material arrives again. More realistic assumptions incorporated in our model provide a new dimension in the area of queueing systems.

Our main objective of this paper is to develop an M\(^{X}/H_{\text{v}}/1\) queueing model with an un-reliable server under multiple vacation policy. Further, we intend to determine the approximate results for the steady-state probability distributions of the queue length using the maximum entropy approach. The rest of the paper is organized as follows. In the ‘Model description’ section, we describe the model and construct the steady-state equations governing the model. Afterwards, in the ‘Probability generating function’ section, we obtain the queue size distribution by using the probability generating function technique. In the ‘Performance measures’ section, we derive various performance indices. The principle of maximum entropy is described in the ‘Maximum entropy principle’ section to establish the approximate results for the expected system size and expected waiting time. In the ‘Numerical illustration and sensitivity analysis’ section, numerical illustrations and sensitivity analysis are presented to validate the
analytical results. Finally, conclusion has been drawn in the ‘Conclusion’ section.

Model description
Consider a single un-reliable, removable server queue with state-dependent rates. We assume that the states of the system are described by the triplet \((i,j,n)\), where \(i=0, 1, ..., k; j = V, B, D\); and \(n = 0, 1, 2,...\). Here \(i = 0\) denotes that the customer is not in service, and \(i = 1, 2,..., k\) denotes that the customer is in the \(i\)th phase service; \(j = V, B, D\) represents that the server is on vacation, busy, and under repair after failure, respectively; and \(n\) denotes the number of customers present in the system. The service time is assumed to follow the \(k\)-type hyper-exponential distribution. It is assumed that \(\mu_i\) (\(i = 1, 2,..., k\)) is the service rate of \(i\)th phase service. Let the probability that the next customer to enter in the service is of type \(i\) be \(q_i\) (\(i = 1, 2,..., k\)) and \(\sum_{i=1}^{k} q_i = 1\).

Other assumptions made to construct the mathematical model are as follows:

- The customers arrive in batches according to the Poisson process with state-dependent arrival rate \(\lambda_j\) given by
  \[
  \lambda_j = \begin{cases} 
  \lambda_0, & \text{if the server is on vacation} \\
  \lambda_1, & \text{if the server is turned on and is in operation} \\
  \lambda_2, & \text{if the server is turned on and is under repair}
  \end{cases}
  \]

  Let \(A\) be the random variable denoting the batch size, and then the batch size distribution is given by
  \[
  c_j = \Pr[A = j], \quad j = 1, 2, ..., d
  \]

  Furthermore, the generating function for the batch size distribution is \(A(z) = \sum_{j=1}^{d} c_j z^j\). It follows that \(E(A) = A'(1)\) and \(E[A(A-1)] = A''(1)\).

- When the breakdown occurs, the server is unable to render service to the customers, but after completing repair provided by a repairman, it works as efficiently as before the failure. The life time and repair time of the server are negative exponentially distributed with mean \(1/\alpha\) and \(1/\beta\), respectively.

- The customers are served according to the first come, first served (FCFS) discipline.

Let us denote the steady-state probabilities depicting the system status as follows:

- \(P_{0,0}(n)\): Probability that there are \(n\) \((n = 0, 1,..., \infty)\) customers in the system and the server is on vacation.
- \(P_{i,0}(n)\): Probability that there are \(n\) \((n = 1, 2,..., \infty)\) customers in the system and the customer in service is in phase \(i\) \((i = 1, 2,..., k)\), when the server is turned on and is in busy state.
- \(P_{i,1}(n)\): Probability that there are \(n\) \((n = 1, 2,..., \infty)\) customers in the system and the customer in service is in phase \(i\) \((i = 1, 2,..., k)\), when the server is turned on and is in a breakdown state.

The steady-state equations governing the model are given as follows:

\[
(\lambda_0 + v)P_{0,V}(n) = \lambda_0 \sum_{k=1}^{n} P_{0,V}(n-k)c_k; \quad n\geq 1 \quad (1)
\]

\[
\lambda_0P_{0,V}(0) = \sum_{j=1}^{k} \mu_jP_{0,V}(1) \quad (2)
\]

\[
(\lambda_1 + \alpha + \mu_i)P_{i,B}(1) = q_i \sum_{j=1}^{k} \mu_jP_{i,B}(2) + \beta P_{i,D}(1); \quad 1 \leq i \leq k \quad (3)
\]

\[
(\lambda_1 + \alpha + \mu_i)P_{i,B}(n) = q_i \sum_{j=1}^{k} \mu_jP_{i,B}(n+1) + \beta P_{i,D}(n) + \lambda_1 \sum_{k=1}^{n-1} P_{i,B}(n-k)c_k; \quad n \geq 2, 1 \leq i \leq k \quad (4)
\]

\[
(\lambda_2 + \beta)P_{i,D}(1) = \alpha P_{i,B}(1); \quad 1 \leq i \leq k \quad (5)
\]

\[
(\lambda_2 + \beta)P_{i,D}(n) = \alpha P_{i,B}(n) + \lambda_1 \sum_{k=1}^{n-1} P_{i,D}(n-k)c_k; \quad n \geq 2, 1 \leq i \leq k \quad (6)
\]

**Probability generating function**

In this section, we present the probability generating function (PGF) technique to obtain the analytical solution of Equations 1 to 6. Let us define the following partial generating functions:

\[
G_{0,V}(z) = \sum_{n=0}^{\infty} z^n P_{0,V}(n), \quad |z| \leq 1 \quad (7)
\]

\[
G_{i,B}(z) = \sum_{n=1}^{\infty} z^n P_{i,B}(n), \quad 1 \leq i \leq k, \quad |z| \leq 1 \quad (8)
\]

\[
G_{i,D}(z) = \sum_{n=1}^{\infty} z^n P_{i,D}(n), \quad 1 \leq i \leq k, \quad |z| \leq 1 \quad (9)
\]
Lemma 1. The expressions for the partial generating functions are obtained as follows:

\[ G_{i,V}(z) = \frac{1}{1 - \rho_v A(z)} P_{0,V}(0) \]  

(10)

\[ G_{i,B}(z) = \frac{N_i(z)}{D(z)} P_{0,V}(0), \quad i = 1, 2, \ldots, k \]  

(11)

\[ G_{i,D}(z) = \frac{\alpha}{(\lambda_2 + \beta - \lambda_2 A(z))} G_{i,B}(z), \quad 1 \leq i \leq k \]  

(12)

where

\[ \rho_v = \frac{\lambda_0}{\lambda_0 + \nu}, \quad D(z) = \prod_{i=1}^{k} \theta_i(z) + \sum_{i=1}^{k} \left[ q_i \mu_i \prod_{j \neq i} \theta_j(z) \right] \]

\[ N_i(z) = \prod_{j=i}^{k} \theta_j(z) q_i \lambda_0 z; \]

\[ \theta_i(z) = \left[ 1 + \lambda A(z) - \left( \lambda_1 + \alpha + \mu_i + \frac{\alpha \beta}{(\lambda C(z) - \lambda_2 \beta)} \right) z \right], \quad i = 1, 2, \ldots, k \]

Proof. For proof, see ‘Proof of Lemma 1’ in the Appendix.

Lemma 2. The probability \( P_{0,V}(0) \) is

\[ P_{0,V}(0) = \frac{1}{1 - \rho_v} + \sum_{i=1}^{k} \rho_v q_i \left( 1 + \frac{\beta}{\lambda_i \beta E[A] + \alpha \lambda_2 E[A]} \right) \frac{1}{\left[ (\lambda_i \beta E[A] + \alpha \lambda_2 E[A]) \sum_{i=1}^{k} \frac{1}{\mu_i} \right]} \]  

(13)

Also, the stability condition is given by

\[ \frac{\alpha}{\beta} < \frac{1 - \lambda_1 E[A]}{\lambda_2 E[A] \sum_{i=1}^{k} \frac{q_i}{\mu_i}} \]  

(14)

Proof. For proof, see ‘Proof of Lemma 2’ in the Appendix.

Theorem 1. The probability generating function of the number of customers in the system is given by

\[ G(z) = \frac{\frac{1}{1 - \rho_v C(z)} + \sum_{i=1}^{k} \lambda_i \mu_i q_i \left( 1 + \frac{\alpha}{(\lambda_2 + \beta - \lambda_2 C(z))} \right) N_i(z)}{D(z)} P_{0,V}(z) \]  

(15)

Proof. For proof, see ‘Proof of Theorem 1’ in the Appendix.

Performance measures

In order to predict the system characteristics under variant circumstances, it is worthwhile to explore key aspects by establishing analytical formulae. In this section, some performance measures in terms of steady-state probabilities are obtained. The long-run probabilities of the server being on vacation, busy, and breakdown are denoted by \( P_{0V} \), \( P_b \), and \( P_d \), respectively. Thus, we obtain

\[ P_{0V} = \sum_{n=1}^{\infty} z^n P_{0,V}(n) = G_{0,V}(1)^{0,V} = \frac{1}{1 - \rho_v} P_{0,V}(0) \]  

(16)

\[ P_b = \sum_{n=1}^{\infty} z^n P_{b,n}(n) = \sum_{i=1}^{k} G_{i,B}(1) \]

\[ = \sum_{i=1}^{k} \frac{\rho_v q_i \left( 1 + \frac{\beta}{\lambda_i \beta E[A] + \alpha \lambda_2 E[A]} \right) \frac{1}{\left[ (\lambda_i \beta E[A] + \alpha \lambda_2 E[A]) \sum_{i=1}^{k} \frac{1}{\mu_i} \right]} P_{0,V}(0)}{\left[ (\lambda_i \beta E[A] + \alpha \lambda_2 E[A]) \sum_{i=1}^{k} \frac{1}{\mu_i} \right]} \]  

(17)

\[ P_d = \sum_{n=1}^{\infty} z^n P_{d,n}(n) = \sum_{i=1}^{k} G_{i,D}(1) \]

\[ = \sum_{i=1}^{k} \frac{\alpha \mu_i q_i \left( 1 + \frac{\beta}{\lambda_i \beta E[A] + \alpha \lambda_2 E[A]} \right) \frac{1}{\left[ (\lambda_i \beta E[A] + \alpha \lambda_2 E[A]) \sum_{i=1}^{k} \frac{1}{\mu_i} \right]} P_{0,V}(0)}{\left[ (\lambda_i \beta E[A] + \alpha \lambda_2 E[A]) \sum_{i=1}^{k} \frac{1}{\mu_i} \right]} \]  

(18)

Theorem 2. The expected number of customers in the system \( (L_N) \) is given by

\[ L_N = G'(1) = \frac{\rho_v E[A]}{(1 - \rho_v)^2} \]

\[ + \sum_{i=1}^{k} \left[ \frac{(\alpha + \beta)(\lambda_i E[A]^{1})(1-N_i(1))2D(1)^{-1}}{\beta D(1)^2} \right] P_{0,V}(0) \]

(19)

where

\[ N_i(1) = \prod_{j=i}^{k} \left( 1 + \frac{\lambda_i \mu_i q_i (1 - \sum_{i=1}^{k} a_i)}{1 - \sum_{i=1}^{k} a_i} \right), \quad i = 1, 2, \ldots, k \]
\[ N_i'(1) = (-1)^i q_i \lambda_0 \left[ b \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} \mu_j ight] \]
\[ + 2 \prod_{i=1}^{k-1} \mu_i \left( k \sum_{i=1}^{k-1} a_i - \sum_{i=1}^{k-1} \prod_{j=1}^{k-1} a_j \right), \quad i = 1, 2, ..., k \]

\[ D'(1) = \prod_{i=1}^{k} \mu_i \sum_{i=1}^{k} a_i \]

subject to the following constraints:

1. Normalizing condition
\[ \sum_{n=0}^{\infty} P_{0,V}(n) + \sum_{n=1}^{k} \sum_{i=1}^{k} P_{i,B}(n) + \sum_{n=1}^{k} \sum_{i=1}^{k} P_{i,D}(n) = 1 \] (21)

2. The probability that the server being busy
\[ \sum_{n=1}^{\infty} \sum_{i=1}^{k} P_{i,B}(n) = \sum_{i=1}^{k} G_{i,B}(1) = \sum_{i=1}^{k} A_i \] (22)

3. The probability that the server is in a breakdown state
\[ \sum_{n=1}^{\infty} \sum_{i=1}^{k} P_{i,D}(n) = \sum_{i=1}^{k} G_{i,D}(1) = \sum_{i=1}^{k} E_i \] (23)

4. The expected number of customers in the system
\[ \sum_{n=0}^{\infty} n P_{0,V}(n) + \sum_{n=1}^{k} \sum_{i=1}^{k} n P_{i,B}(n) + \sum_{n=1}^{k} \sum_{i=1}^{k} n P_{i,D}(n) = L_N \] (24)

where \( A_i = G_{i,B}(1) \), \( E_i = G_{i,D}(1) \), \( 1 \leq i \leq k \), and \( L_N \) are given by Equations 17, 18, and 19, respectively.

After introducing Lagrange’s multipliers corresponding to constraints (21) to (24), we construct Lagrange’s function as

\[
y = -\sum_{n=0}^{\infty} P_{0,V}(n) \log P_{0,V}(n) - \sum_{n=0}^{\infty} \sum_{i=1}^{k} P_{i,B}(n) \log P_{i,B}(n) - \sum_{n=0}^{\infty} \sum_{i=1}^{k} P_{i,D}(n) \log P_{i,D}(n)
\]

(20)

Proof. For proof, see ‘Proof of Theorem 2’ in the Appendix.

Maximum entropy principle

Exact probabilities of the system states of an MX/Hk/1 queueing system with multiple vacations and an unreliable server have not been found earlier to the best information of the authors. In order to evaluate approximate results for the steady-state probabilities, we employ the maximum entropy approach. It is well established that the principle of maximum entropy can be used for estimating probabilistic information measures which is further used to obtain the queue size distribution of various complex queueing systems in different frameworks. In order to obtain the steady-state probabilities \( P_{0,V}(n) \), \( P_{i,B}(n) \), and \( P_{i,D}(n) \) by using the principle of maximum entropy, we formulate the maximum entropy model as follows.

The maximum entropy model

Following El-Affendi and Kouvatlos (1983), the entropy function \( y \) can be mathematically formulated as

\[
y = -\sum_{n=0}^{\infty} P_{0,V}(n) \log P_{0,V}(n) - \sum_{n=0}^{k} \sum_{i=1}^{k} P_{i,B}(n) \log P_{i,B}(n) - \sum_{n=0}^{k} \sum_{i=1}^{k} P_{i,D}(n) \log P_{i,D}(n)
\]

(20)
where \( \theta_i, \eta_i (1 \leq i \leq k) \) and \( \xi_i (1 \leq i \leq k + 1) \) are Lagrange's multipliers corresponding to constraints (21) to (24), respectively.

**The maximum entropy analysis**

The maximum entropy results are obtained by taking the partial derivatives of \( y \) w.r.t. \( P_{0,v}(n), P_{1,b}(n) \), and \( P_{1,d}(n) \) and equating to zero. Thus, we get

\[
\frac{\partial y}{\partial P_{0,v}(n)} = -\log P_{0,v}(n) - 1 - \theta_1 - \xi_{k+1} n = 0
\]

\[
\frac{\partial y}{\partial P_{1,b}(n)} = -\log P_{1,b}(n) - 1 - \theta_1 - \eta_i - \xi_{k+1} n = 0, \quad 1 \leq i \leq k
\]

\[
\frac{\partial y}{\partial P_{1,d}(n)} = -\log P_{1,d}(n) - 1 - \theta_1 - \xi_i - \xi_{k+1} n = 0, \quad 1 \leq i \leq k
\]

From Equations 26 to 28, we get

\[
P_{0,v}(n) = e^{-(1+\theta_1)} e^{-\xi_{k+1} n}, \quad n = 0, 1, \ldots
\]

\[
P_{1,b}(n) = e^{-(1+\theta_1+\eta_i)} e^{-\xi_{k+1} n}, \quad n = 1, 2, \ldots, \quad 1 \leq i \leq k
\]

\[
P_{1,d}(n) = e^{-(1+\theta_1+\xi_i)} e^{-\xi_{k+1} n}, \quad n = 1, 2, \ldots, \quad 1 \leq i \leq k
\]

Denote \( \phi_1 = e^{-(1+\theta_1)}, \psi_i = e^{-\eta_i}, 1 \leq i \leq k \) and \( \delta_i = e^{-\xi_i}, 1 \leq i \leq k+1 \).

Then, Equations 29 to 31 can be written as

\[
P_{0,v}(n) = \phi_1 \phi_{k+1}^{n}, \quad n = 0, 1, \ldots
\]

\[
P_{1,b}(n) = \phi_1 \psi_i \phi_{k+1}^{n}, \quad n = 1, 2, \ldots, \quad 1 \leq i \leq k
\]

\[
P_{1,d}(n) = \phi_1 \phi_i \phi_{k+1}^{n}, \quad n = 1, 2, \ldots, \quad 1 \leq i \leq k
\]

On substituting the values of \( P_{0,v}(n), P_{1,b}(n), \) and \( P_{1,d}(n) \) from Equations 32 to 34 into Equations 21 to 23, we obtain

\[
\sum_{n=0}^{\infty} \phi_1 \phi_{k+1}^{n} = \frac{\phi_1}{1-\phi_{k+1}} = 1 - \sum_{i=1}^{k} (A_i + E_i)
\]

\[
\sum_{n=1}^{\infty} \phi_1 \psi_i \phi_{k+1}^{n} = \frac{\phi_1 \psi_i \phi_{k+1}^{n}}{1-\phi_{k+1}} = A_i, \quad 1 \leq i \leq k
\]

It follows from Equations 35 to 37 that

\[
\phi_1 = (1-\rho)(1-\delta_{k+1})
\]

\[
\psi_i = \frac{A_i}{(1-\rho)\delta_{k+1}}, \quad 1 \leq i \leq k
\]

\[
\delta_i = \frac{E_i}{(1-\rho)\delta_{k+1}}, \quad 1 \leq i \leq k
\]

where \( \rho = \sum_{i=1}^{k} (A_i + E_i) \).

On substituting the values of \( \phi_1, \psi_i \), and \( \delta_i (1 \leq i \leq k) \) from Equations 38 to 40 into Equation 24 and after doing some algebraic manipulations, we obtain

\[
\delta_{k+1} = \frac{L_N-\rho}{1+L_N-\rho}
\]

On substituting the values of \( \phi_1, \psi_i \), and \( \delta_i \) from Equations 38 to 40 into Equations 32 to 34 and using Equation 40, we finally get

\[
P_{0,v}(n) = \left( \frac{1-\rho}{1+L_N-\rho} \right) \left( \frac{L_N-\rho}{1+L_N-\rho} \right)^n, \quad n = 0, 1, \ldots
\]

\[
P_{1,b}(n) = \left( \frac{A_i}{1+L_N-\rho} \right) \left( \frac{L_N-\rho}{1+L_N-\rho} \right)^{n-1}, \quad n = 1, 2, \ldots, \quad 1 \leq i \leq k
\]

\[
P_{1,d}(n) = \left( \frac{E_i}{1+L_N-\rho} \right) \left( \frac{L_N-\rho}{1+L_N-\rho} \right)^{n-1}, \quad n = 1, 2, \ldots, \quad 1 \leq i \leq k
\]

**The expected waiting time in the system**

Let \( W_s \) and \( \bar{W}_s \) denote the exact and the expected waiting time in the system, respectively. Then,

\[
W_s = \frac{L_N}{\lambda_{eff}}
\]

where \( \lambda_{eff} = [\lambda_0 P_V + \lambda_1 P_B + \lambda_2 P_D] E[A] \)
Following the work of Wang et al. (2007), the approximate expected waiting time in the system is given by the approximate expected waiting time in

\[
W_S = \frac{1}{\nu} \left[ \sum_{i=1}^{k} \left( \sum_{j=1}^{\infty} \frac{nq_i}{\mu_j} + \frac{1}{\nu} + \frac{q_i}{2\mu_i} \left( \frac{E[A]}{E[A]} - 1 \right) \right) P_{0,V}(n) + \sum_{i=1}^{k} \left( \sum_{j=1}^{\infty} \frac{nq_i}{\mu_j} + \frac{1}{\nu} + \frac{q_i}{2\mu_i} \left( \frac{E[A]}{E[A]} - 1 \right) \right) P_{L,V}(n) \right]
\]

Substituting the values of \(P_{0,V}(n), P_{L,V}(n)\) and \(P_{L,D}(n)\) from Equations 42 to 44 into Equation 46, the approximate expected waiting time in the system is given by

\[
W_S = \frac{1}{\nu} \left[ \sum_{i=1}^{k} A_i \left( \sum_{i=1}^{\infty} E_i \right) + \sum_{i=1}^{k} q_i \left( L_N + \frac{1}{2} \left( \frac{E[A]}{E[A]} - 1 \right) \right) + \sum_{i=1}^{k} E_i \right] \frac{1}{\beta}
\]

**Numerical illustration and sensitivity analysis**

In this section, we present a numerical simulation by taking the illustration of production of heat transfer equipment (HTE) discussed in the ‘Background’ section. For developing the code of computational program, we have used the ‘MATLAB’ software. For computation purposes, we assume that the raw materials arrive in batches of fixed batch size \(k = 3\). The arrival rates are chosen as \(\lambda_1 = 0.5\), \(\lambda_2 = 0.9\), and \(\lambda_3 = 0.7\). The service times of the machine when producing these equipment are \(\mu_1 = 1\), \(\mu_2 = 2\), and \(\mu_3 = 3\). The processing of the equipment may be interrupted with rate \(\alpha = 0.8\) and again becomes available for processing with rate \(\beta = 2\). Further, the server may go for multiple vacations with rate \(\nu = 0.09\). The expected number of these equipment in the system is obtained by using Equations 1 to 19 as \(L_N = 36\).

Now, we present the numerical results to demonstrate the effects of different parameters on various performance indices. The accuracy of numerical results is examined by comparing the exact waiting time (\(W_S\)) obtained in the ‘Maximum entropy principle’ section using the probability generating function approach with the approximate waiting time (\(\tilde{W}_S\)) obtained by the maximum entropy principle (MEP) of the \(M^{\infty}/H_k/1\) queuing system under multiple vacation policy. Relative percentage error is tabulated for this purpose. The variations of different parameters on the average queue length are shown in Figures 1 and 2 graphically.

Table 1 summarizes the numerical results for long-run fraction of time of the server being in different states by varying the parameters \(\lambda, \mu, \alpha, \beta\), and \(\nu\) for two different cases of \(q_i\) (case 1: \((q_1, q_2, q_3) = (0.6, 0.3, 0.1)\) and case 2: \((q_1, q_2, q_3) = (0.7, 0.2, 0.1)\)). For the sake of convenience, we choose the default parameters \(\lambda_1 = 0.9, \lambda_2 = 0.8, \lambda_3 = 0.7, \mu_1 = 1, \mu_2 = 2, \mu_3 = 3, \alpha = 0.2, \beta = 1\), and \(\nu = 0.9\). It is noticed that \(P_V\) shows a decreasing trend with respect to the increasing values of \(\lambda, \alpha, \) and \(\nu\), but an increasing trend has been found with other parameters for both cases. Similarly, \(P_B\) and \(P_D\) increase as we increase the values of \(\lambda, \alpha, \) and \(\nu\) and decrease with increasing values of \(\mu\) and \(\beta\) for both cases.

Figure 1a,b,c depicts the effect of different parameters on average queue length for various sets of heterogeneous arrival rate \((\lambda_0 = 1.8\lambda, \lambda_1 = 0.5\lambda, \lambda_2 = 0.6\lambda)\) shown by discrete lines and homogeneous arrival rate \((\lambda_0 = \lambda_1 = \lambda_2 = \lambda)\) shown by continuous lines. We observe that the average queue length is higher for the heterogeneous arrival rate in comparison to the homogeneous arrival rate on increasing the breakdown rate of the server. The queue length shows a gradual decreasing trend on increasing the repair rate and the vacation rate. Further, Figure 2a,b,c visualizes the effect of batch size on the average queue length. It is observed that the average queue length reveals an increasing trend with increasing values of \(\alpha\) and \(\beta\) while shows a decreasing trend with increasing values of \(\nu\). The tractability of numerical results shows that our model can be easily implemented for the quantitative assessment of the performance of many real-time congestion systems.

**Methods**

In this paper, an \(M^X/H_k/1\) queue under multiple vacations and an un-reliable server with varying arrival rates is studied. The probability generating function technique is used to determine various performance measures in explicit form. Then, MEP is further employed to compare the approximate results with exact results. For validating the analytical results, the sensitivity analysis is also carried out.
Figure 1 Average queue length vs. (a) $\alpha$, (b) $\beta$, and (c) $\nu$ for homogeneous and heterogeneous arrival rates.

Figure 2 Average queue length vs. (a) $\alpha$, (b) $\beta$, and (c) $\nu$ for different batch sizes.
The role of queueing analysis based on MEP lies in the fact that it helps system designers and managers to take decisions based on the performance indices determined using the probability distribution of the system size in terms of available information by MEP approach.

The numerical illustration presented demonstrates that maximum entropy analysis is a simple approach to deal with complex scenarios for real-life congestion situations and can be easily applied to complex queueing scenarios for which performance measures are not easily obtained by using a classical approach. Based on the numerical experiment and sensitivity analysis carried out, we overall conclude that the comparative analysis of approximate results with exact results has demonstrated that the results obtained by MEP are reasonably good. The average queue length is higher for heterogeneous arrival rate in comparison to homogeneous arrival rate. The trends are more perceptible for larger batch size, which is quite obvious as the congestion increases significantly if the batch size of arriving customers is large.

### Table 1 Effect of various parameters on long-run fraction of time of the server being in different states

| Case 1: \((q_1, q_2, q_3) = (0.6, 0.3, 0.1)\) | Case 2: \((q_1, q_2, q_3) = (0.7, 0.2, 0.1)\) |
|---|---|
| \(\lambda\) 0.700 | 0.900 |
| 0.725 | 0.925 |
| 0.750 | 0.940 |
| 0.775 | 0.955 |
| 0.800 | 0.970 |
| \(\mu\) 1.000 | 1.000 |
| 1.050 | 1.050 |
| 1.100 | 1.100 |
| 1.150 | 1.150 |
| 1.200 | 1.200 |
| \(a\) 0.100 | 0.100 |
| 0.200 | 0.200 |
| 0.300 | 0.300 |
| 0.400 | 0.400 |
| 0.500 | 0.500 |
| \(\beta\) 1.000 | 1.000 |
| 1.500 | 1.500 |
| 1.750 | 1.750 |
| 2.000 | 2.000 |
| \(v\) 0.100 | 0.100 |
| 0.200 | 0.200 |
| 0.300 | 0.300 |
| 0.400 | 0.400 |
| 0.500 | 0.500 |

### Table 2 Comparison between the exact and the approximate results of waiting times

| Case 1: \((q_1, q_2, q_3) = (0.6, 0.3, 0.1)\) | Case 2: \((q_1, q_2, q_3) = (0.7, 0.2, 0.1)\) |
|---|---|
| \(\lambda\) 0.600 | 0.800 |
| 0.700 | 0.900 |
| 0.800 | 0.900 |
| \(\mu\) 1.000 | 1.000 |
| 1.250 | 1.250 |
| \(a\) 0.100 | 0.100 |
| 0.200 | 0.200 |
| 0.300 | 0.300 |
| 0.400 | 0.400 |
| \(\beta\) 1.000 | 1.000 |
| 2.000 | 2.000 |
| \(v\) 0.010 | 0.010 |
| 0.012 | 0.012 |

### Conclusion

In this paper, an \(M^X/H_k/1\) queue with multiple vacations and an unreliable server has been studied in order to facilitate various performance indices in explicit form by using an analytical approach based on the generating function method and maximum entropy principle. The incorporation of some more realistic features such as
Again put \( i = 2 \) in Equation 50 and using Equation 52 into Equation 50, we obtain
\[
q_i q_i G_{i,B}(z) + \left[ \lambda_i z A(z) - \left( \lambda_i + \alpha + \mu_i \right) z + q_i \mu_i \right] G_{i,B}(z) + q_i \sum_{j=1}^{k-1} \mu_j G_{i,B}(z) = q_i z A P_{0,V}(0)
\]
\[ (54) \]

Similarly, repeating this process for \( i = k \), we get
\[
q_k \sum_{j=1}^{k-1} \mu_j G_{i,B}(z) + \left[ \lambda_i z A(z) - \left( \lambda_i + \alpha + \mu_i \right) z + q_i \mu_i \right] G_{i,B}(z) = q_i z A P_{0,V}(0)
\]
\[ (55) \]

We use Cramer’s rule to solve Equations 53 to 55. Now we get
\[
G_{i,B}(z) = \frac{N_i(z)}{D(z)} P_{0,V}(0), \quad i = 1, 2, \ldots, k
\]
\[ (56) \]

**Proof of Lemma 2**

Using Lemma 1, we obtain
\[
G_{0,V}(1) = \lim_{z \to 1} G_{0,V}(z) = \frac{1}{1-\rho_v} P_{0,V}(0)
\]
\[ (57) \]
\[
G_{i,B}(1) = \lim_{z \to 1} G_{i,B}(z) = \frac{\rho_i \left[ (1-\lambda_i B(z) + \alpha_0 E(z) \sum_{i=1}^{k-1} \mu_i \right] P_{0,V}(0)}{(\lambda_i B(z) + \alpha_0 E(z) \sum_{i=1}^{k} \mu_i \beta)}
\]
\[ (58) \]
\[
G_{i,D}(1) = \lim_{z \to 1} G_{i,D}(z) = \frac{\alpha}{\beta} G_{i,B}(1), \quad i = 1, 2, \ldots, k
\]
\[ (59) \]

where \( \rho_i = \frac{\lambda_i}{\mu_i} \).

The L-Hospital rule has been applied to compute the above results.

To determine \( P_{0,V}(0) \), we use the normalizing condition given by
\[
G(1) = G_{0,V}(1) + \sum_{i=1}^{k} \left( G_{i,B}(1) + G_{i,D}(1) \right)
\]
\[ (60) \]

On substituting the values of \( G_{0,V}(1) \), \( G_{i,B}(1) \), and \( G_{i,D}(1) \) from Equations 57 to 59 into Equation 60, we obtain the value of \( P_{0,V}(0) \).
For finding the result of the stable condition, we use the condition
\[ 0 < P_{0,V}(0) < 1. \] (61)

Using Equation 13 into Equation 61, we have
\[ 0 < (\lambda_1 \beta + \alpha \lambda_2) E[A] \sum_{i=1}^{k} \frac{q_i}{\mu_i} - \beta < 1. \] (62)

After some algebraic manipulations, Equation 62 provides the result given in Equation 14.

**Proof of Theorem 1**

In order to prove Equation 15, we have
\[
G(z) = \sum_{n=0}^{\infty} z^n P_{0,V}(n) + \sum_{n=1}^{\infty} \sum_{i=1}^{k} z^n P_{i,D}(n) \\
+ \sum_{n=1}^{\infty} \sum_{i=1}^{k} z^n P_{i,D}(n) \\
= G_{0,V}(z) + \sum_{i=1}^{k} [G_{i,B}(z) + G_{i,D}(z)]
\] (63)

On substituting the values of \( G_{0,V}(z) \), \( G_{i,B}(z) \), and \( G_{i,D}(z) \) from Lemma 1 into Equation 63, we get Equation 15.

**Proof of Theorem 2**

The average system size is computed using
\[
L_N = \lim_{z \to 1} G(z).
\]

The L-Hospital rule is applied twice to compute the results given in Equation 19.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

MJ has worked on \( M/G/H_1 \) queueing system under multiple vacation policy. Various performance indices including queue size using generating function and the approximate formulas for waiting time of the customers using the maximum entropy principle have been obtained by RS. GC has performed numerical results by taking an illustration.

**Acknowledgements**

The authors are thankful to the learned referee and editor for their valuable comments and suggestions for the improvement of the paper.

**Author details**

1 Department of Mathematics, IIT, Roorkee 247667, India. 2 Department of Mathematics, JK Lakshmipat University, Jaipur 302026, India. 3 Department of Mathematics, I.B.S., Khandari, Agra, 282002, India.

Received: 8 December 2012 Accepted: 13 November 2013 Published: 05 Dec 2013

**References**

Annamalathan R, Jayakumar S (2004) Analysis of bulk queue with multiple vacations and closedown times. Int. J Inform Manage Sci 15(1):45–60
Baba Y (1986) On the \( M/G/1 \) queue with vacation time. Oper Res Lett 5:93–98
Choudhury G, Deka K (2008) An \( M/G/1 \) retrial queueing system with two phases of service subject to server breakdown and repair. Perf Eval 65(10):714–724
Choudhury G, Tadig J (2009) An \( M/G/1 \) queue with two phases of service subject to the server breakdown and delayed repair. Appl Math Comp 330:2699–2709
El-Affendi MA, Kovatsos DD (1983) A maximum entropy analysis of the \( M/G/1 \) and \( G/M/1 \) queueing systems at equilibrium. Acta Inf 19:339–355
Jain M, Dhadak MU (2003) Queueing distribution \( M^p/G/1 \) model using principle of maximum entropy. J Dec Math Sci 8(1):33–44
Jain M, Sharma GC, Sharma R (2012) Maximum entropy approach for un-reliable server queueing model with optional bulk service. JKAU: Eng Sci 23(2):89–113
Jain M, Sharma GC, Sharma R (2013) Unreliable server \( M/G/1 \) queue with multioptional services and multioptional vacations. Int J Math Oper Res 5(2):145–169
Jain M, Singh M (2000) Diffusion process for optimal flow control of \( G/G/c \) finite capacity queue. J MACT 33:88–88
Ke JC, Chang FM (2009) \( M^p//G_1//G_2//1 \) retrial queue under Bernoulli vacation schedules with general repeated attempts and starting failures. Appl Math Model 33(4):443–458
Ke JC (2005) Modified T vacation policy for an \( M/G/1 \) queueing system with unreliable server and startup. Appl Math Model 41:1267–1277
Ke JC (2007) Operating characteristic analysis on the \( M^p/G/1 \) system with variant vacation policy and balking. Appl Math Model 31(7):1321–1337
Ke JC, Lin CH (2006) Maximum entropy solutions for batch-arrival with an unreliable server and delaying vacations. Appl Math Comp 183(2):1328–1340
Ke JC, Lin CH (2008) Maximum entropy approach for batch-arrival queue under \( N \) policy with an unreliable server and single vacation. Comp Appl Math 22(1):1–15
Ke JC, Lin CH, Yang JY, Zhang ZG (2009) Optimal (d, c) vacation policy for a finite buffer \( M/G/C \) queue with unreliable servers and repairs. Appl Math Model 33(10):3949–3962
Kumar BK, Kochwesari SP (2005) An \( M/M/2 \) queueing system with heterogeneous servers and multiple vacations. Math Comp Model 41(13):1415
Li H, Tan NS, Ma ZY (2007) Performance analysis of \( G/M/1 \) queue with working vacations. Oper Res Lett 35:595–600
Ono Y, Glick SV (2008) Maximum entropy analysis of the \( M^p//M^q//1 \) queueing system with multiple vacation and server breakdowns. Comp Indus Eng 54(4):1078–1086
Sharma R (2010) Threshold \( N \)-policy for \( M^p/H_1^p//1 \) queueing system with un-reliable server and vacations. J Int Acad Phy Sci 14:1–11
Singh CJ, Jain M, Kumar B (2012) Analysis of \( M/G/1 \) queueing model with state dependent arrival and vacation. J Indus Eng Int 8(2):1–18
Singh CJ, Jain M, Kumar B (2013) Analysis of unreliable bulk queue with state dependent arrival. J Indus Eng Int 9(1):1–9
Tan N, Zhang ZG (2002) The discrete-time \( G//Geo//1 \) queue with multiple vacations. Queue Sys 40(3):283–294
Wang J (2004) An \( M/G/1 \) queue with second optional service and server breakdowns. Comp Math Appl 47(10–11):1713–1723
Wang KH, Chan MC, Ke JC (2007) Maximum entropy analysis of \( M^p//M^q//1 \) queueing system with multiple vacations and server breakdowns. Comp Indus Eng 52(1–2):192–202
Wang KH, Chang KW, Swaladan BD (1999) Optimal control of a removable and non-reliable in an infinite and a finite \( M/H_1^p//1 \) queueing system. Appl Math Model 23:651–666
Wang KH, Huang KB (2009) A maximum entropy approach for the \( p, N \)-policy \( M/G/1 \) queue with unreliable server and single vacation. Appl Math Model 33(4):2024–2034
Wang KH, Kao HT, Chen G (2004) Optimal management of a removable and non-reliable server in an infinite and a finite \( M/H_1^p//1 \) queueing system. Qual Tech Quant Mang 12(2):325–339
Wang WL, Xu GQ (2009) The well-posedness of an \( M/G/1 \) queue with second optional service and server breakdown. Comp Math Appl 57(5–6):729–739
Wu DA, Takagi H (2006) An \( M/G/1 \) queues with multiple working vacations and exhaustive service discipline. Perf Eval 63(7):654–681