State Dependent $M/H_{k}/1/N$ Queueing System With Service Interruption

Madhu Jain$^1$ and Deepa Chauhan$^2$*

1 Department of Mathematics, Indian Institute of Technology, Roorkee, India.
2 Department of Mathematics, Allenhouse Institute of Technology, Kanpur, India.

Abstract: The objective of this paper is to analyze $M/H_{k}/1$ queueing system with setup and unreliable server for both finite and infinite capacity under N-policy. The customers arrive to the system in accordance with Poisson process. The service times follow the $k$-type hyper exponential distribution. The breakdown and repair times are assumed to follow a negative exponential distribution. The arriving customers may balk, depending upon the number of customers present in the system. The generating function method is used to derive queue size distribution of the number of the customers present in the system. The cost function has been derived in term of cost elements related to different situations in order to determine the optimal operating policy. To explore the effects of various parameters on cost and other indices, the sensitivity analysis is carried out by taking numerical illustrations.

MSC: 60J05.

Keywords: N-policy, Unreliable server, $M/H_{k}/1$ queue, Balk, Set up, Loss and delay, Queue size, Generating function.

1. Introduction

In this investigation $M/H_{k}/1$ queueing system with service interruption is considered for infinite capacity where removable and non-reliable server operates under N-policy. The term removable server may be interpreted as turning on and turning off the server in the system depending on the number of units waiting in the system. The non-reliable server means that the server is subject to unpredictable breakdowns. Under N-policy, the server turned on when $N \geq 1$ or more customers are present in the system and turned off when system is empty. [3] studied the operating characteristics of a batch arrival queueing system under N policy. In many realistic situations, the server takes some times before starting the service of the customers that is known as set up or start up time of the server. [1] studied a queueing system where the service station operates under an N-policy with early setup. Queueing model with service interruptions is helpful in predicting the performance of various machining systems. [2] examined a multi-component machining system consisting of m operating machines along with k type of spares under N-policy. In this paper we worked on optimal management of a removable and non-reliable server for $M/H_{k}/1$ queueing system by incorporating the startup time of the server and balking behavior of the customers. In section 2, we outline some assumptions and notations in order to construct the mathematical model of the system. Section 3 contains the equations governing the model and analytical expressions of the probability generating functions for different states. In next section 4, average queue length for infinite model has been obtained. Some performance characteristics and cost function are also derived. Finally conclusions are drawn in section 5.

* E-mail: drdeepachauhan07@gmail.com
2. Model Description

Consider $M/H_k/1$ queue with unreliable server, balking and setup under N-policy along with following assumptions:

1. Two types of customers called delay and loss customers join the system according to Poisson process with rate $\lambda_1$ and $\lambda_2$, respectively. When the system is empty, the integrated rate of arriving customers is assumed to be $\lambda_0$. The delay customer may balk on finding the server in broken down state with probability $\upsilon$.

2. The server turns on when $N$ customers are accumulated in the system and initiates the service of the customers after a setup time which is exponentially distributed with rate $\theta$. The server turns off when system becomes empty.

3. $(i,j,n)$ denotes the state of the system with $n = 0, 1, 2, \ldots$; $i = 0, 1, 2, \ldots, k$; $j = F, D, W$; where $n$ is the number of customers in the system, $i = 0$ denotes that there is no customer in the service and I denotes that the customer is in the service of type $i$ ($i = 1, 2, \ldots, k$). Also $j$ denotes the status of the server as

   $$ j = \begin{cases} 
   F, & \text{server is turned off and in idle state} \\
   W, & \text{server is turned on and in operating state} \\
   D, & \text{server is turned on and in breakdown state} 
   \end{cases} $$

4. Whenever the server is turned on and being in $i^{th}$ ($i = 1, 2, \ldots, k$) phase of service, it may breakdown at any time with rate $\alpha_i$ ($i = 1, 2, \ldots, k$); the inter-failure time of the server follows a negative exponential distribution.

5. Whenever the server fails during the $i^{th}$ phase service, it immediately goes for repair and repair time is assumed to be negative exponentially distributed with repair rate $\beta_i$ ($i = 1, 2, \ldots, k$).

6. For finite capacity model, $L$ is the capacity of the system.

7. The switch over from repair to working state of the server is instantaneous.

Notations

- $N$ Threshold level of queue length when server turns on
- $\lambda_0$ Integrated arrival rate of the customers when server is idle
- $\lambda_1$ Arrival rate of delay customers
- $\lambda_2$ Arrival rate of loss customers
- $\mu_i$ Service rate of the $i^{th}$ type service, $i = 1, 2, \ldots, k$
- $\alpha_i$ Breakdown rate of the server during $i^{th}$ phase service, $i = 1, 2, \ldots, k$
- $\beta_i$ Repair rate of the server
- $\upsilon$ Joining probability of the customer
- $\theta$ Startup rate of the server
- $P_{0,F}(n)$ The probability of being $n$ ($n = 0, 1, 2, \ldots, k$) customers in the system when the server is turned off.
- $P_{i,W}(n)$ The probability of being $n$ ($n = 0, 1, 2, \ldots, k$) customers in the system when the server is turned on and the customer is in service of type $i$ ($i = 1, 2, 3, \ldots, k$).
- $P_{i,D}(n)$ The probability of being $n$ ($n = 0, 1, 2, \ldots, k$) customers in the system and the server is broken-down state while providing the service to the customer of type $i$ ($i = 1, 2, 3, \ldots, k$).
- $H_i(z)$ The probability generating function of $P_i(n)$
- $E(N_i)$ Expected number of customers in the system when the server is in the state $i$. 
3. Infinite Capacity Queue

The steady state equations for N-policy M/H/1 queueing system with infinite capacity are given by:

\[
\lambda_0 P_{0,F}(0) = \sum_{i=1}^{k} \mu_i P_{i,w}(1), \quad n = 0
\]  
(1)

\[
\lambda_0 P_{0,F}(n) = \lambda_0 P_{0,F}(n-1), \quad 1 \leq n \leq N - 1
\]  
(2)

\[
(\lambda_0 + \theta) P_{0,F}(n) = \lambda_0 P_{0,F}(n-1), \quad n \geq N
\]  
(3)

\[
(\lambda_1 + \lambda_2 + \mu_i + \alpha_i) P_{i,w}(1) = q_i \left[ \sum_{j=1}^{k} \mu_j P_{j,w}(2) \right] + \beta_i P_{i,d}(1), \quad i = 1, 2, ..., k
\]  
(4)

\[
(\lambda_1 + \lambda_2 + \mu_i + \alpha_i) P_{i,w}(n) = (\lambda_1 + \lambda_2) P_{i,w}(n-1) + q_i \left[ \sum_{j=1}^{K} \mu_j P_{j,w}(n+1) \right] + \beta_i P_{i,d}(n), \quad i = 1, 2, ..., k
\]  
(5)

\[
(\lambda_1 + \lambda_2 + \mu_i + \alpha_i) P_{i,w}(N) = (\lambda_1 + \lambda_2) P_{i,w}(N-1) + q_i \left[ \sum_{j=1}^{k} \mu_j P_{j,w}(N+1) \right]
\]  

\[+ \beta_i P_{i,d}(N) + q_i \lambda_0 P_{0,F}(N-1) + \theta P_{0,F}(N), \quad i = 1, 2, ..., k
\]  
(6)

\[
(\lambda_1 + \lambda_2 + \mu_i + \alpha_i) P_{i,w}(n) = (\lambda_1 + \lambda_2) P_{i,w}(n-1) + q_i \left[ \sum_{j=1}^{k} \mu_j P_{j,w}(n+1) \right]
\]  

\[+ \beta_i P_{i,d}(n) + \theta P_{0,F}(n), \quad i = 1, 2, ..., k, \quad n \geq N + 1
\]  
(7)

\[
(\lambda_1 \theta + \beta_i) P_{i,d}(1) = \alpha_i P_{i,w}(1)
\]  
(8)

\[
(\lambda_1 \theta + \beta_i) P_{i,d}(n) = \alpha_i P_{i,w}(n) + \lambda_2 \theta P_{i,d}(n-1), \quad n \geq 2
\]  
(9)

3.1. Probability Generating Functions

For obtaining the closed form expression of \( P_{0,F}(0) \) we use the probability generating function technique. Let us define the following probability generating functions:

\[
G_{0,F}(Z) = \sum_{n=0}^{N-1} Z^n P_{0,F}(n)
\]  
(10)

\[
G_{i,w}(Z) = \sum_{n=i}^{\infty} Z^n P_{i,w}(n)
\]  
(11)

\[
G_{i,d}(Z) = \sum_{n=i}^{\infty} Z^n P_{i,d}(n)
\]  
(12)

Using (1), (2) and (10), we get

\[
G_{0,F}(Z) = \left( 1 + \sum_{n=0}^{N-1} Z^n \right) P_{0,F}(0) = 1 - \frac{Z^N}{1-Z} P_{0,F}(0)
\]  
(13)

Multiplying (3)-(7) by appropriate power of \( z \) and summing over \( n \), we obtain

\[
(\lambda_1 + \lambda_2 + \mu_i + \alpha_i) Z G_{i,w}(Z) + q_i \lambda_0 Z P_{0,F}(0) = q_i \left[ \sum_{j=1}^{k} \mu_j G_{j,w}(Z) \right] + (\lambda_1 + \lambda_2) Z^2 G_{j,w}(Z)
\]  

\[+ \beta_i Z G_{i,d}(Z) + q_i \lambda_0 Z^{n+1} P_{0,F}(0), \quad i = 1, 2, ..., k
\]  
(14)

Now multiplying (8)-(9) by appropriate power of \( Z \) and summing over \( n \), we get

\[
(\lambda_1 \theta Z - \lambda_1 \theta - \beta_i) G_{i,d}(Z) + \alpha_i G_{i,w}(Z) = 0
\]  
(15)
We determine \( P \) by using normalizing equation
\[
G_{i,w}(Z) + q_1 \left\{ \sum_{j=2}^{k} \mu_j G_{j,w}(Z) \right\} = q_1 \lambda_0 Z P_{0,F}(0) - q_1 \lambda_0 Z^{N+1} P_{0,F}(0)
\] (16)

Similarly for \( i = 2 \), we get
\[
q_2 \mu_2 G_{1,w}(Z) + \left\{ (\lambda_1 + \lambda_2) Z^2 - \left( \lambda_1 + \lambda_2 + \mu_2 + \alpha_2 + \frac{\alpha_2 \beta_2}{\lambda_1 \varphi - \lambda_1 \varphi - \beta_2} \right) Z \right\} G_{1,w}(Z) = q_2 \lambda_0 Z P_{0,F}(0) - q_2 \lambda_0 Z^{N+1} P_{0,F}(0)
\] (17)

Repeating this process for \( i = k \), we get
\[
q_k \left\{ \sum_{j=1}^{k-1} \mu_j G_{j,w}(Z) \right\} + q_k \mu_k G_{k,w}(Z) + \left\{ (\lambda_1 + \lambda_2) Z^2 - \left( \lambda_1 + \lambda_2 + \mu_k + \alpha_k + \frac{\alpha_k \beta_k}{\lambda_1 \varphi - \lambda_1 \varphi - \beta_k} \right) Z \right\} G_{k,w}(Z) = q_k \lambda_0 Z P_{0,F}(0) - q_k \lambda_0 Z^{N+1} P_{0,F}(0)
\] (18)

Using Cramer’s rule we solve (16)-(18) and obtain probability generating functions
\[
G_{i,w}(Z) = \frac{N_i(Z)}{D(Z)} P_{0,F}(0), \quad i = 1, 2, ..., k
\] (19)
\[
G_{i,d}(Z) = \frac{\alpha_i N_i(Z)}{\left( \lambda_1 \varphi Z - \lambda_1 \varphi - \beta_i \right) D(Z)} P_{0,F}(0), \quad i = 1, 2, ..., k,
\] where
\[
N_i(Z) = \prod_{j \neq i} \phi_j(Z) q_i \lambda_0 Z (1 - Z^N) P_{0,F}(0), \quad i = 1, 2, ..., k
\]
\[
D(Z) = \prod_{i=1}^{k} \phi_i(Z) + \sum_{i=1}^{k} q_i \mu_i \prod_{j \neq i} \phi_j(Z)
\]
\[
\phi_i(Z) = (\lambda_1 + \lambda_2) Z^2 - \left( \lambda_1 + \lambda_2 + \mu_i + \alpha_i + \frac{\alpha_i \beta_i}{\lambda_1 \varphi Z - \lambda_1 \varphi - \beta_i} \right) Z
\]

Let \( G(Z) \) denote the p.g.f. of the number of customers in the system, then
\[
G(Z) = \sum_{n=0}^{N-1} Z^n P_{0,F}(0) + \sum_{n=1}^{\infty} Z^n \sum_{i=1}^{k} \{ P_{i,w}(n) + P_{i,d}(n) \}
\]
\[
= \sum_{i=1}^{k} \{ G_{i,w}(Z) + G_{i,d}(Z) \}
\] (21)

To evaluate \( G_{0,F}(1) \), \( G_{1,w}(1) \) and \( G_{1,d}(1) \), we apply L’Hospital rule in equations (13), (19) and (20) respectively, and get
\[
G_{0,f}(1) = N P_{0,F}(0)
\] (22)
\[
G_{1,w}(1) = \frac{N \lambda_0 \prod_{i=1}^{k} \frac{2n_i}{\mu_i} \left( \lambda_1 + \lambda_2 \right) \sum_{i=1}^{k} \frac{2n_i}{\mu_i} + \lambda_1 \varphi \sum_{i=1}^{k} \frac{2n_i}{\mu_i \beta_i} - 1 + \lambda_1 \varphi}{\left( \lambda_1 + \lambda_2 \right) \sum_{i=1}^{k} \frac{2n_i}{\mu_i} + \lambda_1 \varphi \sum_{i=1}^{k} \frac{2n_i}{\mu_i \beta_i} - 1 + \lambda_1 \varphi} P_{0,F}(0), \quad i = 1, 2, ..., k
\] (23)

By using normalizing equation
\[
G(1) = G_{0,F}(1) + \sum_{i=1}^{k} \{ G_{i,w}(1) + G_{i,d}(1) \} = 1
\] (24)

We determine \( P_{0,F}(0) \) as
\[
P_{0,F}(0) = \frac{\left( \lambda_1 + \lambda_2 \right) \sum_{i=1}^{k} \frac{2n_i}{\mu_i} + \lambda_1 \varphi \sum_{i=1}^{k} \frac{2n_i}{\mu_i \beta_i} - 1}{\left( \lambda_1 + \lambda_2 \right) \sum_{i=1}^{k} \frac{2n_i}{\mu_i} + \lambda_1 \varphi \sum_{i=1}^{k} \frac{2n_i}{\mu_i \beta_i} - 1 + \lambda_1 \varphi} N
\] (25)
4. Average Queue Length

The expected number of customers in the N-policy \(M/H_k/1\) queueing system with non-reliable server is obtained as

\[
L_1 = \frac{N(N - 1)}{2} + \sum_{i=1}^{k} \left\{ (\alpha_i + \beta_i)(N_i''(1)D_i'(1) - N_i'(1)D_i''(1)) \right\} P_{b,r}(0)
\]

where

\[
N_i'(1) = \prod_{j \neq i} \mu_j \mu_0 N, \quad i = 1, 2, ..., k
\]

\[
N_i''(1) = \prod_{i=1}^{k} q_i \mu_0 N \left\{ 2(\lambda_1 + \lambda_2) + (N + 1) \prod_{j \neq i} \mu_j + 2\lambda_1 \theta \sum_{j \neq i} \frac{\alpha_j}{\beta_j} \right\}, \quad i = 1, 2, ..., k
\]

\[
D_i'(1) = \prod_{i=1}^{k} \mu_i \left\{ 1 - (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{q_i}{\mu_i} + (N + 1) \prod_{j \neq i} \mu_j + 2\lambda_1 \theta \sum_{j \neq i} \frac{\alpha_j}{\beta_j} \right\}, \quad i = 1, 2, ..., k
\]

\[
D_i''(1) = \prod_{i=1}^{k} (-2\mu_i) \left[ (\lambda_1 + \lambda_2) \left( \sum_{i=1}^{k} \frac{q_i}{\mu_i} + \lambda_1 \theta \sum_{i=1}^{k} \frac{\alpha_i}{\beta_i} \right) \right] + (\lambda_1 \theta)^2 \left( \sum_{i=1}^{k} \frac{q_i \alpha_i}{\beta_i^2 \mu_i^2} + \prod_{i=1}^{k} \frac{\alpha_i}{\beta_i} \right) + (\lambda_1 + \lambda_2)^2 \right], \quad i = 1, 2, ..., k
\]

In order to design optimal policy we shall derive some performance indices for different system states, which are defined as follows:

1. Idle period (I): This is the length of time per cycle during which the server is turned off.

2. Busy period \((B_i)\): This is the length of time per cycle when server is turned on and is operational and type \(i\) \((i = 1, 2, ..., k)\) phase service to the customer is being provided.

3. Breakdown period (D): This is the length of time per cycle during which the server is found to be broken down and under repair.

Let the expected length of idle period, the busy period of type \(i\) \((i = 1, 2, ..., k)\), the breakdown period and the busy cycle be denoted by \(E[I]\), \(E[B_i]\), \(E[D]\) and \(E[C]\), respectively. The expected busy cycle is the sum of idle period, busy period, breakdown period and is given by

\[
E[C] = E[I] + \sum_{i=1}^{k} E[B_i] + E[D]
\]

\[
= E[I] + E[B] + E[D], \quad \text{where} \quad \sum_{i=1}^{k} E[B_i] = E[B]
\]

According the memory less property, the length of idle period is the sum of \(N\) exponential random variables each having mean \(1/\lambda\), therefore \(E[I] = N/\lambda\).

4.1. Performance Characteristics

The long run fraction time for which the server is idle, busy and broken down respectively, are given by

\[
P_F = \frac{E[I]}{E[C]} = \frac{\left[ (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{q_i}{\mu_i} + \lambda_1 \theta \sum_{i=1}^{k} \frac{\alpha_i}{\mu_i \beta_i} - 1 \right]}{\left[ (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{q_i \alpha_i}{\mu_i \beta_i^2} + \lambda_1 \theta \sum_{i=1}^{k} \frac{\alpha_i}{\mu_i} + (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{\alpha_i}{\beta_i} - 1 \right] + \lambda_0 \sum_{i=1}^{k} \frac{\alpha_i}{\mu_i} + \lambda_0 \prod_{i=1}^{k} \frac{\alpha_i}{\mu_i \beta_i}}
\]
The above inequalities can be shown to reduce to

\begin{equation}
P_W = \frac{E[B]}{E[C]} = N\lambda_0 \prod_{i} \frac{q_i}{\mu_i}
\end{equation}

\begin{equation}
P_D = \frac{E[D]}{E[C]} = N\lambda_0 \prod_{i} \frac{q_i}{\mu_i}
\end{equation}

Using eqs (27) - (30), the number of busy cycles per unit time is given by:

\begin{equation}
\frac{1}{E[C]} = \frac{\lambda \left( (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{q_i}{\mu_i} + \lambda_2 \sum_{i=1}^{k} \frac{q_i}{\mu_i} - 1 \right)}{\left( (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{q_i}{\mu_i} + \lambda_1 \theta \sum_{i=1}^{k} \frac{q_i}{\mu_i} - 1 \right)}
\end{equation}

### 4.2. Cost Function

Now we develop a cost model by considering N as a decision variable for M/H_\alpha/1 queueing system with infinite capacity. Our objective is to determine the optimal value of N (say N*) which minimizes the expected total cost. In order to construct cost function, the following cost elements are used:

- **C_{su}** Start-up cost per unit time when the server is turned on
- **C_{sd}** Shutdown cost per unit time when the server is turned off
- **C_h** Holding cost per unit time per customer present in the system
- **C_o** Cost per unit time for keeping the server on
- **C_f** Cost per unit time for keeping the server off
- **C_b** Breakdown cost per unit time for the failed server

The expected total cost function per unit time is given by

\begin{equation}
T[C(N)] = C_h L_1 + C_f E[I] + C_D E[B] + C_o E[D] + (C_{su} + C_{sd}) \frac{1}{E[C]}
= C_h \frac{N-1}{2} + (C_{su} + C_{sd}) \frac{\lambda \left( (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{q_i}{\mu_i} + \lambda_1 \theta \sum_{i=1}^{k} \frac{q_i}{\mu_i} - 1 \right)}{\left( (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{q_i}{\mu_i} + \lambda_1 \theta \sum_{i=1}^{k} \frac{q_i}{\mu_i} - 1 \right)}
\end{equation}

### 4.3. The Optimal Operating Policy

The following inequality can be used to determine the optimal threshold parameter (i.e. N*) that minimize the expected total cost

\begin{equation}
T[C(N* + 1)] \geq T[C(N*)] \leq T[C(N* - 1)]
\end{equation}

The above inequalities can be shown to reduce to

\begin{equation}
N^*_1(N^*_1 - 1) \leq (C_{su} + C_{sd}) \frac{2\lambda \left( (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{q_i}{\mu_i} + \lambda_2 \sum_{i=1}^{k} \frac{q_i}{\mu_i} - 1 \right)}{\left( (\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{q_i}{\mu_i} + \lambda_1 \theta \sum_{i=1}^{k} \frac{q_i}{\mu_i} - 1 \right)} C_h
\leq N^*_1(N^*_1 + 1)
\end{equation}
Assuming \( N \) as a continuous variable we obtain the approximate optimal value of \( N^* \) by setting \( \frac{dT(C(N+1))}{dN} = 0 \). Hence the optimal value of \( N \) is approximately given by

\[
N^*_1 = \sqrt{\frac{(C_{su} + C_{sd})}{(\lambda_1 + \lambda_2) \sum_{i=1}^{k} \frac{\mu_i}{\rho_i} + \lambda_1 \theta \sum_{i=1}^{k} \frac{a_i}{\mu_i \rho_i} - 1}}
\]

(35)

5. Concluding Remarks

In this paper, \( M/H_k/1 \) model for infinite capacity with balking is developed. The optimal value of threshold parameter is determined which is helpful to minimize the expected total operating cost. The incorporation of state dependent rates, startup time and balking probability makes our model more versatile from application point of view. The concept of unreliable server has been included, which can be realized in many real time systems namely computer systems, production and manufacturing systems, etc..

References

