TWMS J. App. and Eng. Math. V.15, N.xx, 2025, pp. 2269-2283.

# ANALYSIS OF SECOND OPTIONAL SERVICE SYSTEM WITH A COLD STANDBY SERVER THAT IS RELIANT ON THE SYSTEM SIZE

P. VIJAYA LAXMI[1*], G. ANJALIDEVI[1], §

Abstract. In this article, we examine a second optional service queueing system using two types of servers, viz. the main operating server and a reliable standby server. All arriving customers receive the first essential service (FES), and only a few may thereafter request a second optional service (SOS) with some probability. During FES and SOS services, the primary operational server may break down. The server is promptly sent for repair if a break down arises and the standby server will be replaced only if the system size is $q$ $(\geq 1)$; otherwise, customers would queue up while the main server is being repaired and resumes the service. We also derive the necessary and sufficient condition for the system to be stable. The model's steady state solution is discovered using the matrix geometric approach. Further, multiple system performance measures are obtained and a cost optimization problem is taken into consideration. Graphs are used to display the numerical outcomes.

Keyword: Queue, Standby server, Matrix geometric method, Breakdown and Repairs.

AMS Subject Classification: 60K25, 68M20, 90B22.

## 1. Introduction

Second optional service (SOS) systems have been the focus of extensive research in queueing theory for a long time. Particularly, these queues find applications in hospitals, educational system, banking, transport, manufacturing systems, etc. For example, consider the airport management system, where it carries out major responsibilities such as passenger processing, arrival and departure operations, baggage tagging and handling, information dissemination, luggage trolley, bus services etc. Further, the in-flight meal, commonly known as airline cuisine, restaurants, free internet, ATMs, vehicle rentals, currency exchange, etc., are few of the services that airports carry out in addition to their primary duties. Madan [1] first looked into an $M/G/1$ queueing system with SOS, where some customers would require an SOS right away once the first essential service (FES) is finished. Later Jain and Chauhan [2] made an assessment of utilizing the matrix geometric approach to operate the vacation queue with the SOS and an unreliable server. Vijaya

[1] Department of Applied Mathematics, Andhra University, Visakhapatnam, Andhra Pradesh, India.
e-mail: vijayalaxmiau@gmail.com; ORCID: https://orcid.org/0000-0002-8721-5148.
e-mail: anjalidevigilaka@gmail.com; ORCID: https://orcid.org/0009-0005-1475-7105.
[*] Corresponding author.

§ Manuscript received: July 26, 2024; accepted: October 21, 2024.
TWMS Journal of Applied and Engineering Mathematics, Vol.15, No.9; 

Laxmi et al. [3] analyzed correlated reneging in SOS with working vacations in a Markovian queue. Recently, Vijaya Laxmi et al. [4] contemplated SOS and feedback equipped Markovian batch service queue in both steady and transition states by using the method of probability generating functions. A few interesting papers include, [5], [6], [7], and the references therein.

Server breakdown and repair are prominent features in queueing theory. The SOS system has received more attention recently for both breakdowns and repairs. Due to their broad usefulness for call centres, communication and computers systems, as well as manufacturing and production technologies, these systems are extensively investigated. Gray et al. [8] explored a queueing system with breakdowns and vacations. Later, Tarabia [9] contemplated on an $M/M/1$ queue's transient and steady state analysis including balking, catastrophes, server breakdowns and repairs. Recently Seenivasan et al. [10] made an assignment on state-dependent customers in an $M/M/1$ queue, a failure server, a single working vacation, and feedback. For some recent papers on $M/M/1$ queue with server breakdown and repair, we cite [11], [12], [13], [14], [15] etc.

Standby server is another factor that is generally taken into account in queueing systems. A queueing system might encounter a sudden failure, causing service to be halted until the server is repaired. In this case, when a customer's service is interrupted standby server might be needed. There is an extensive literature on standby server that has been studied in various forms by numerous authors. Khalaf et al. [16] explored on a batch arrival queueing system with a standby server during vacations or when the primary server is undergoing repair. In this case, the standby server only serves customers during the time the main server is away on vacation or when the main server is down for repairs due to an unexpected failure from time to time. Chakravarthy and Kulshrestha [12] studied a queueing system that accounts for server failures, repair, vacations, and standby server. Ayyappan and Karpagam [17] first dealt with a bulk queue with an unreliable server, instant feedback, $N$-policy, several Bernoulli vacation schedules, and a backup server. Then, Ayyappan and Karpagam [18] investigated a bulk service queue with a standby server, several vacations, overloading, and an unreliable server. Recently, Vijaya Laxmi and Bhavani [19] made an assessment on both warm and cold standbys in a repairable SOS queueing system. In general, there are three kinds of standby servers available, viz., warm-standby, hot-standby, and cold-standby servers. When a standby server's failure rate is non-zero and lower than that of the primary operating server, it is termed as warm. A server is called hot standby server if its failure rate is equal to that of the primary operating server, and it is referred to as cold standby server when it has no failures. We have employed a cold standby server for this article, as it avoids further delays in services due to breakdown of main operating server.

Many authors have employed the matrix geometric technique in their works. EL-Rayes et al. [20] studied through the use of matrix geometric techniques, an infinite stochastic process algebra models. Lakshmi and Ramanath [21] investigated an $M/M/1$ two-phase multi-optional retrial queue with Bernoulli feedback, impatient users, and a server that is prone to failure and repair is solved by using the matrix geometric technique. Joshi et al. [22] made an assessment on method using matrix geometric method to investigate the $M/M/1$ model that is being repaired. Matrix geometric technique was studied for performance evaluation of multiple service systems by Shah et al. [23]. A few interesting papers include Jain and Chauhan [2], [19], [12], [3] etc.

The behaviour of an $M/M/1$ queue in steady state is examined in this paper with two types of servers (main server and cold standby server), two types of services (FES and SOS), breakdown and repair. In this article, a new form of standby server has been developed and examined. When a certain threshold value for system length is reached, the standby server is replaced by the main server, and continue the service. If the system length falls below the threshold value, then the remaining customers will have to wait for the main server to resume the service.

The following is how the remaining paper is set up: Practical application of the model is given in Section 2. Model description can be found in Section 3. In Section 4, the mathematical formulation, equations for steady states and the matrix geometric solution are provided. Performance indicators are listed in Section 5. Section 6 includes the cost analysis and results of numerical analyses. Finally, conclusions are presented in Section 7.

## 2. Practical Application of the Proposed Model

Consider a call center of a major telecommunications company that provides customer support for a variety of services including billing, technical issues, and account management. This call center operates with a primary group of highly trained agents (main server) who handle a wide range of customer inquiries. These agents are proficient in resolving basic customer service queries (FES) such as billing questions, account updates, and general service information. However, some customer issues require more specialized knowledge, such as resolving complex technical problems or handling high priority cases (SOS). These specialized tasks are typically handled by the same group of primary agents.

Breakdown scenario: Imagine a scenario where a sudden technical disrupts the main server's operations, rendering the primary group of highly trained agents unavailable. This could be due to issues like a system outage, network failure, or other technical difficulties that prevent the agents from accessing the necessary tools and information to assist customers.

Standby server activation: To ensure continuity of service, the call center has a secondary group of less experienced agents or automated systems (standby server) on standby. These secondary agents are capable of handling basic inquiries but may not be as adept at resolving more complex issues. The standby server is activated only when the call volume exceeds a certain threshold ($q$), ensuring that the call center's resources are utilized efficiently. When the primary group of agents becomes unavailable, the standby server is immediately activated if the number of incoming calls exceeds the predetermined threshold. Customers who need basic support are directed to the secondary agents, while more complex issues are queued until the primary agents are back online. This system ensures that basic customer service continues uninterrupted, minimizing wait times and maintaining a satisfactory level of service.

For example, during a system outage, customers calling with simple billing inquiries or requests for account information are quickly solve by the standby server. Meanwhile, those with more complex technical issues are informed of the delay and their concerns will be addressed once the primary system is restored. This approach not only helps manage customer expectations but also maintains a level of operational efficiency even during disruptions. This model allows the call center to maintain service levels and customer satisfaction by efficiently balancing the workload between primary and standby servers. By implementing such a queuing system, the company can ensure that basic customer service remains uninterrupted during technical difficulties, thereby reducing customer frustration and maintaining trust in the company's support services.

## 3. Model Information

Consider an endless single server queue with a cold standby server that is introduced based on the system size and a second alternative service system. Arrivals take place via

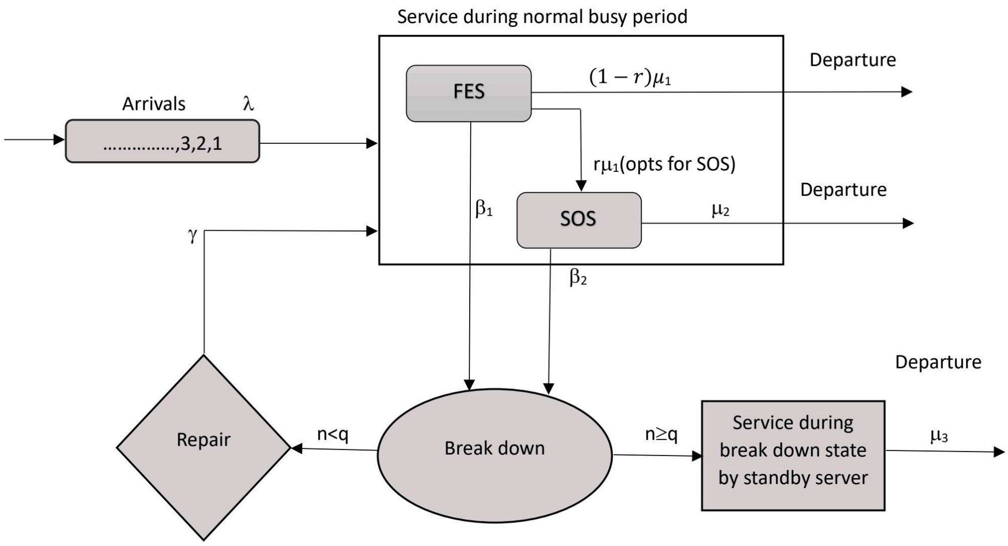


Figure 1. General structure of the model

the Poisson process with arrival rate $\lambda$. All arriving customers receive FES from the main server and after FES, customers may choose to leave the system with probability $(1-r)$ or request SOS with probability $r$. The service times during FES and SOS are distributed exponentially with rates $\mu_1$ and $\mu_2$, respectively. Main server may break down during busy period of FES or SOS and immediately it is sent for repair. Failure times during FES and SOS are distributed exponentially with rates $\beta_1$ and $\beta_2$, respectively and repair time is also distributed exponentially with rate $\gamma$. The standby server will become operational only if the number of customers in the system reaches a particular threshold value $q$, $q(\geq 1)$. However, when the system size is smaller than $q$ and main server is broken down, those customers must have to wait in the system until the main server resumes service.

The following cases are explained below when the main server is in break down state.

- Case (i) : The system size is smaller than $q$.
  Customers in this situation will have to wait for the main server to resume service because the standby server would not be provided.
- Case (ii): The system size exceeds or is equal to $q$.
  In this case, a standby server will continue to provide the service and is deactivated whenever the the system size goes down the threshold value $q$.
- Case (iii): It may be noted that during the break down state of the main server (FES or SOS), the standby server will be providing service with a service rate $\mu_3$, which is generally lower than the main server's service rates. It is assumed that the standby server provides only one type of service and after completion of service, the customers depart from the system.

Figure 2 depicts the state transition diagram for the model under consideration.

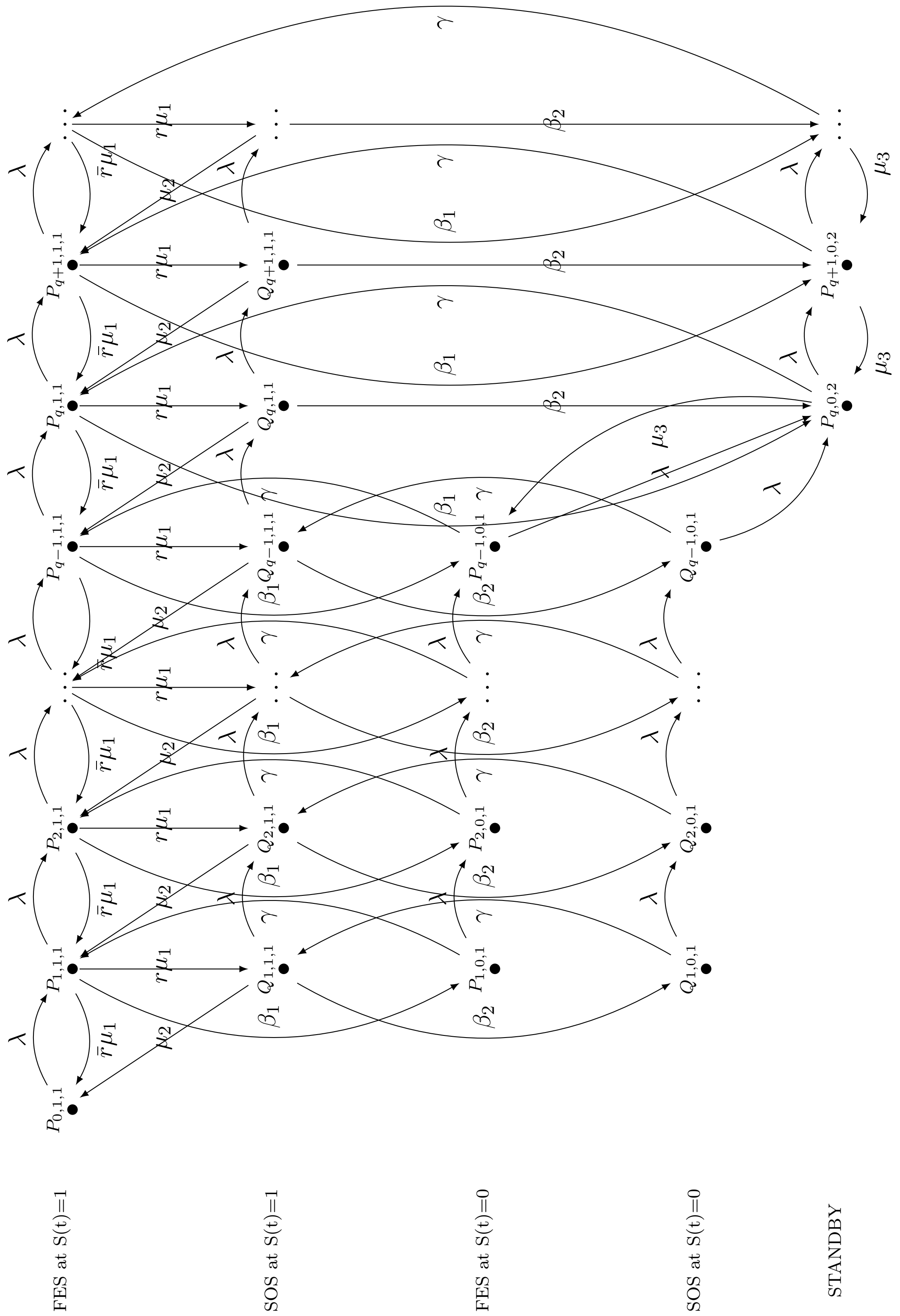


FIGURE 2. State transition diagram
where, $\bar{r} = (1 - r)$.

## 4. Mathematical Formulation and Matrix Geometric Solution of the Model

Let $N(t)$ represent the system's total number of customers. $S(t)$ represent the server's current state, where

$$S(t)=\begin{cases} 0, & \text{server is in breakdown state,} \\ 1, & \text{server is in normal busy state.} \end{cases}$$

and $K(t)$ be the type of the service given by

$$K(t)=\begin{cases} 1, & \text{standby server is not replaced,} \\ 2, & \text{standby server is replaced.} \end{cases}$$

Evidently, this system defined by the process $\{N(t), S(t), K(t), t \geq 0\}$ forms a continuous-time Markov process with state space $\chi = \{(n, s, k) : n \geq 0,\ s = 0, 1,\ k = 1, 2\}$.
Let $P_{n,1,1}$ represent the probability that there are $n$ customers in the system, the main server in normal busy period and renders FES, $n \geq 0$.
Let $Q_{n,1,1}$ represent the probability that there are $n$ customers in the system, the main server in normal busy period and render SOS, $n \geq 1$.
The probability that there are $n$ customers in the system, the main server is broken down during FES, and the standby server is not be replaced is given by $P_{n,0,1}$, $0 \leq n \leq q-1$.
The probability that there are $n$ customers in the system, the main server is broken down during SOS and standby server is not replaced is given by $Q_{n,0,1}$, $1 \leq n \leq q-1$.
The probability that there are $n$ number of consumers in the system, the main server in break down state, and service provided by standby server is given by $P_{n,0,2}$, $n \geq q$.

### 4.1. Steady State Probabilities.

The following Kolmogorov forward difference equations are written in steady state:

1. Main server is in normal busy period and providing FES

$$\lambda P_{0,1,1} = (1-r)\mu_1 P_{1,1,1} + \mu_2 Q_{1,1,1}, \tag{1}$$

$$(\lambda + \mu_1 + \beta_1) P_{n,1,1} = \lambda P_{n-1,1,1} + (1-r)\mu_1 P_{n+1,1,1} + \mu_2 Q_{n+1,1,1} + \gamma P_{n,0,1}, 1 \leq n \leq q-1, \tag{2}$$

$$(\lambda + \mu_1 + \beta_1) P_{n,1,1} = \lambda P_{n-1,1,1} + (1-r)\mu_1 P_{n+1,1,1} + \mu_2 Q_{n+1,1,1} + \gamma P_{n,0,2}, n \geq q. \tag{3}$$

2. Main server is in normal busy period and providing SOS

$$(\lambda + \mu_2 + \beta_2) Q_{1,1,1} = r\mu_1 P_{1,1,1} + \gamma Q_{1,0,1}, \tag{4}$$

$$(\lambda + \mu_2 + \beta_2) Q_{n,1,1} = r\mu_1 P_{n,1,1} + \gamma Q_{n,0,1} + \lambda Q_{n-1,1,1}, 2 \leq n \leq q-1, \tag{5}$$

$$(\lambda + \mu_2 + \beta_2) Q_{n,1,1} = r\mu_1 P_{n,1,1} + \lambda Q_{n-1,1,1}, n \geq q. \tag{6}$$

3. Main server is broken down during FES and standby server is not replaced

$$(\lambda + \gamma) P_{1,0,1} = \beta_1 P_{1,1,1}, \tag{7}$$

$$(\lambda + \gamma) P_{n,0,1} = \beta_1 P_{n,1,1} + \lambda P_{n-1,0,1}, 2 \leq n \leq q-2, \tag{8}$$

$$(\lambda + \gamma) P_{n,0,1} = \beta_1 P_{n,1,1} + \lambda P_{n-1,0,1} + \mu_3 P_{n+1,0,2}, n = q-1. \tag{9}$$

4. Main server is broken down during SOS and standby server is not replaced

$$(\lambda + \gamma) Q_{1,0,1} = \beta_2 Q_{1,1,1}, \tag{10}$$

$$(\lambda + \gamma) Q_{n,0,1} = \beta_2 Q_{n,1,1} + \lambda Q_{n-1,0,1}, \leq n \leq q-1. \tag{11}$$

5.The service is replaced by a backup server after the main server breaks down.

$$(\lambda+\gamma+\mu_3)P_{q,0,2} = \lambda P_{q-1,0,1}+\mu_3 P_{q+1,0,2}+\lambda Q_{q-1,0,1}+\beta_1 P_{q,1,1}+\beta_2 Q_{q,1,1}, \quad (12)$$
$$(\lambda+\gamma+\mu_3)P_{n,0,2} = \lambda P_{n-1,0,2}+\mu_3 P_{n+1,0,2}+\beta_1 P_{n,1,1}+\beta_2 Q_{n,1,1}, n\geq q+1. \quad (13)$$

4.2. **Matrix Geometric Method.** Now, in order to ascertain the stationary probabilities, we employ the matrix geometric method. According to Neuts [24], The process's infinitesimal generator $Q$ could be represented as follows using the system of equations (1) to (13):

$$\mathbf{Q}=\begin{pmatrix} \hat{\mathbf{A}}_0 & \hat{\mathbf{C}}_0 \\ \hat{\mathbf{B}}_1 & \hat{\mathbf{A}}_1 & \hat{\mathbf{C}}_1 \\ & \hat{\mathbf{B}}_2 & \hat{\mathbf{A}}_2 & \hat{\mathbf{C}}_2 \\ & & \ddots & \ddots & \ddots \\ & & & \hat{\mathbf{B}}_{q-1} & \hat{\mathbf{A}}_{q-1} & \hat{\mathbf{C}}_{q-1} \\ & & & & \hat{\mathbf{B}}_q & \hat{\mathbf{A}}_q & \hat{\mathbf{C}}_q \\ & & & & & \hat{\mathbf{B}}_{q+1} & \hat{\mathbf{A}}_q & \hat{\mathbf{C}}_q \\ & & & & & & \hat{\mathbf{B}}_{q+1} & \hat{\mathbf{A}}_q & \hat{\mathbf{C}}_q \\ & & & & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

where $\hat{\mathbf{A}}_0=\begin{pmatrix} -\lambda \end{pmatrix}$, $\hat{\mathbf{C}}_0=\begin{pmatrix} -\lambda & 0 & 0 & 0 \end{pmatrix}$, $\hat{\mathbf{B}}_1=\begin{pmatrix} (1-r)\mu_1 \\ \mu_2 \\ 0 \\ 0 \end{pmatrix}$,

$$\hat{\mathbf{A}}_i=\begin{pmatrix} -(\lambda+\mu_1+\beta_1) & r\mu_1 & \beta_1 & 0 \\ 0 & -(\lambda+\mu_2+\beta_2) & 0 & \beta_2 \\ \gamma & 0 & -(\lambda+\gamma) & 0 \\ 0 & \gamma & 0 & -(\lambda+\gamma) \end{pmatrix}, 1\leq i\leq q-1,$$

$$\hat{\mathbf{B}}_i=\begin{pmatrix} (1-r)\mu_1 & 0 & 0 & 0 \\ \mu_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, 2\leq i\leq q-1,$$

$$\hat{\mathbf{B}}_q=\begin{pmatrix} (1-r)\mu_1 & 0 & 0 & 0 \\ \mu_2 & 0 & 0 & 0 \\ 0 & 0 & \mu_3 & 0 \end{pmatrix}, \hat{\mathbf{B}}_i=\begin{pmatrix} (1-r)\mu_1 & 0 & 0 \\ \mu_2 & 0 & 0 \\ 0 & 0 & \mu_3 \end{pmatrix}, i\geq q+1,$$

$$\hat{\mathbf{A}}_i=\begin{pmatrix} -(\lambda+\mu_1+\beta_1) & r\mu_1 & \beta_1 \\ 0 & -(\lambda+\mu_2+\beta_2) & \beta_2 \\ \gamma & 0 & -(\lambda+\gamma+\mu_3) \end{pmatrix}, i\geq q,$$

$$\hat{\mathbf{C}}_i=\begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix}, 1\leq i\leq q-2, \hat{\mathbf{C}}_{q-1}=\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \\ 0 & 0 & \lambda \end{pmatrix}, \hat{\mathbf{C}}_i=\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}, i\geq q.$$

4.2.1. *Stability Condition.* The topic of stability is of utmost importance in queue theory particularly dealing with the infinite buffer queues. As there is a possibility of system overflow or instability, one has to impose some conditions on the parameters. Below we derive a stability condition which will be used for the systems stability.

**Theorem 4.1.** *The necessary and sufficient condition for the system to be stable is that*
$$\frac{\lambda[\gamma(\mu_2+\beta_2)+r\mu_1(\gamma+\beta_2)+\beta_1(\mu_2+\beta_2)]}{\gamma\mu_1\mu_2+\gamma\beta_2(1-r)\mu_1+\mu_3[r\mu_1\beta_2+\beta_1(\mu_2+\beta_2)]}<1.$$

*Proof.* Let us define the matrix $\xi = \hat{\mathbf{B}}_{q+1} + \hat{\mathbf{A}}_q + \hat{\mathbf{C}}_{\mathbf{q}}$ given by
$$\xi = \begin{pmatrix} -(r\mu_1+\beta_1) & r\mu_1 & \beta_1 \\ \mu_2 & -(\mu_2+\beta_2) & \beta_2 \\ \gamma & 0 & -\gamma \end{pmatrix}$$
there exists a stationary probability vector $\mathbf{Y}= (\mathrm{y}_0, \mathrm{y}_1, \mathrm{y}_2)$ of $\xi$ such that
$$\mathbf{Y}\xi = \mathbf{0} \quad and \quad \mathbf{Ye_3} = \mathbf{1}, \tag{14}$$
where $\mathbf{e_3} = [1,1,1]^T$. Generally, $\mathbf{e}_n$ is the column vector of dimension $n$ with each element set to one. By using Neuts [24], the necessary and sufficient condition for the stability is as follows:
$$\mathbf{Y\hat{C}_q e_3} < \mathbf{Y\hat{B}_{q+1} e_3}. \tag{15}$$
Solving equations (14) and (15), we get
$$(1-r)\mu_1 y_0 + \mu_2 y_2 + \mu_3 y_3 > \lambda(y_0+y_1+y_2). \tag{16}$$
$$\Longrightarrow \frac{\lambda[\gamma(\mu_2+\beta_2)+r\mu_1(\gamma+\beta_2)+\beta_1(\mu_2+\beta_2)]}{\gamma\mu_1\mu_2+\gamma\beta_2(1-r)\mu_1+\mu_3[r\mu_1\beta_2+\beta_1(\mu_2+\beta_2)]}<1 \tag{17}$$
where
$$\begin{aligned} y_0 &= \frac{\gamma(\mu_2+\beta_2)}{\omega},\ y_1 = \frac{\gamma r\mu_1}{\omega},\ y_2 = \frac{\gamma\mu_1\beta_2+\beta_1(\mu_2+\beta_2)}{\omega}, \\ \omega &= \gamma(\mu_2+\beta_2)+r\mu_1(\gamma+\beta_2)+\beta_1(\mu_2+\beta_2) \end{aligned}$$
□

4.2.2. *Steady State Solution.* Let $\mathbf{P}$ represent the stationary probability vector of the generator $\mathbf{Q}$, where $\mathbf{0}$ is the row vector with all elements set to zero and $\mathbf{e}_n$ is the column vector of dimension $n$ with each element set to one. The vector $\mathbf{P}$ is divided as $\mathbf{P} = [\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2, ...]$, where $\mathbf{P}_0 = [P_{0,1,1}]$,
$\mathbf{P}_n = [P_{n,1,1}, Q_{n,1,1}, P_{n,0,1}, Q_{n,0,1}]$, $1 \le n \le q-1$,
$\mathbf{P}_n = [P_{n,1,1}, Q_{n,1,1}, P_{n,0,2}]$, $n \ge q$.
Clearly, the sub-vectors of $\mathbf{P}$ corresponding to different levels satisfy the following equations when the stability condition is satisfied.
$$\mathbf{P}_n = \mathbf{P}_{q+1}\mathbf{R}^{n-(q+1)},\ n > q+1, \tag{18}$$
where $\mathbf{R}$ is the minimal non-negative solution of the matrix quadratic equation represented as
$$\mathbf{C}_q + \mathbf{R}\mathbf{A}_q + \mathbf{R}^2\mathbf{B}_{q+1} = \mathbf{0}. \tag{19}$$
Indeed, the Quasi-birth–death process can only be positive recurring if and only if the spectral radius $Sp(\mathbf{R}) < 1$. Neuts [24] created an iterative approach for computing $\mathbf{R}$ numerically. We determine the successive approximation using the starting iteration $\mathbf{R}_0 = 0$, and calculate the successive approximations using the recurrence relation
$$\mathbf{R}_{n+1} = -(\hat{\mathbf{C}}_q + \mathbf{R}_n^2\hat{\mathbf{B}}_{q+1})\hat{\mathbf{A}}_q^{-1}, n \ge 0.$$

The sequence $\{\mathbf{R}_n\}$ is non-decreasing and converges monotonically to the rate matrix $\mathbf{R}$. We terminate the iteration and return with the solution $\mathbf{R}$ when $\| \mathbf{R}_{n+1} - \mathbf{R}_n \| < \epsilon$, where $\epsilon$ is tolerance error.
The governing set of differential equations is represented by equation $\mathbf{PQ} = \mathbf{0}$ which is expanded to generate the following set of equations

$$\mathbf{P}_0\hat{\mathbf{A}}_0 + \mathbf{P}_1\hat{\mathbf{B}}_1 = 0, \tag{20}$$

$$\mathbf{P}_{n-1}\hat{\mathbf{C}}_{n-1} + \mathbf{P}_n\hat{\mathbf{A}}_n + \mathbf{P}_{n+1}\hat{\mathbf{B}}_{n+1} = 0, 1 \leq n \leq q, \tag{21}$$

$$\mathbf{P}_{n-1}\hat{\mathbf{C}}_q + \mathbf{P}_n\hat{\mathbf{A}}_q + \mathbf{P}_{n+1}\hat{\mathbf{B}}_{q+1} = 0, n \geq q+1, \tag{22}$$

and the normalizing condition

$$\sum_{n=0}^{\infty} \mathbf{P}_n\mathbf{e}_n = 1. \tag{23}$$

From equations (20) to (22), after some mathematical manipulations, we get

$$\mathbf{P}_{n-1} = \mathbf{P}_n\phi_n, \ 1 \leq n \leq q+1, \tag{24}$$

$$\mathbf{P}_{q+1}[\phi_{q+1}\hat{\mathbf{C}}_q + \hat{\mathbf{A}}_q + \mathbf{R}\hat{\mathbf{B}}_{q+1}] = \mathbf{0}, \tag{25}$$

where

$$\phi_1 = -\hat{\mathbf{B}}_1(\hat{\mathbf{A}}_0^{-1}), \ \phi_n = -\hat{\mathbf{B}}_n(\hat{\mathbf{A}}_{n-1} + \phi_{n-1}\mathbf{C}_{n-2})^{-1}, \ 1 \leq n \leq q+1.$$

Using equations (23) and (24), we obtain

$$\mathbf{P}_{q+1}\left[\sum_{l=1}^{q+1}\prod_{i=l}^{q+1}\phi_i + (\mathbf{I} - \mathbf{R})^{-1}\right]\mathbf{e}_n = 1. \tag{26}$$

Solving equations (25) and (26), yields $\mathbf{P}_{q+1}$. The remaining steady state probabilities can be obtained recursively in terms of $\mathbf{P}_{q+1}$ using (18).

## 5. Performance Measures

We provide several performance measures in this part that can reflect the behaviour of the model for slight changes in the parameters.

- The likelihood that the backup server will replace the primary server is

$$P_{ms} = \sum_{n=q}^{\infty} P_{n,0,2} + \sum_{n=1}^{q-1} Q_{n,0,1}$$

- the likelihood of the main server getting repaired is

$$P_r = \sum_{n=1}^{q-1} P_{n,0,1} + \sum_{n=1}^{q-1} Q_{n,0,1}$$

- The likelihood that the system being empty is

$$P_0 = P_{0,1,1}$$

- Expected system size when the main server is in normal busy periods of both FES and SOS only is

$$E[L_b] = \sum_{n=1}^{\infty} nP_{n,1,1} + \sum_{n=1}^{\infty} nQ_{n,1,1}$$

- When the service is delivered through a backup server only, the expected system size is

$$E[L_s] = \sum_{n=q}^{\infty} nP_{n,0,2} + \sum_{n=1}^{q-1} nQ_{n,0,1}$$

- Total expected system size regardless of the state of the server is given by

$$E[L] = E[L_b] + E[L_s]$$

- The expected system size is when the primary server fails without replacing the backup server.

$$L_{smb} = \sum_{n=1}^{q-1} nP_{n,0,1} + \sum_{n=1}^{q-1} nQ_{n,0,1}$$

## 6. Cost Analysis and Numerical Investigations

6.1. **Cost Analysis.** In this section, we construct a function of total estimated cost per unit of time based on system parameters. It is crucial to minimize the cost as much as possible by determining the optimal service rate. For the concerned queueing system, we define the cost factors relative to the main activities, denoting the cost per unit of time as follows:
$c_1 \equiv$ while the main server is typically busy,
$c_2 \equiv$ when a standby server is used in place of the main server,
$c_3 \equiv$ when a server is idle,
$c_4 \equiv$ the main server is being repaired,
$c_5 \equiv$ the fixed cost per unit of time for the standby server.

Using these cost definitions, the total expected cost function per unit of time is given by:

$$f[\mu_1] = c_1 E[L_b] + c_2 q E[L_s] + c_3 \mu_1 \mu_2 P_0 + c_4 Pr + c_5 E[L_{smb}].$$

The cost minimization problem can then be expressed as:

$$f[\mu_1] = \underset{\mu_1}{\text{minimize}}\, f[\mu_1].$$

The minimization of the total cost function depends on the performance measures and system parameters. The cost components $E[L_b]$, $E[L_s]$, $P_r$, $E[L_{smb}]$ and $P_0$ reflect the probabilistic behavior of the system and directly affect the overall cost function based on system utilization and repair rates.

6.2. **Numerical Investigation.** Graphs are used to illustrate various numerical examples in this section. The model's parameters are presumed to be $\lambda = 1.0$, $\mu_1 = 6.0$, $\mu_2 = 3.0$, $\mu_3 = 2.0$, r = 0.6, $\gamma = 0.8$, $\beta_1 = 0.7$, $\beta_2 = 0.4$, q = 6, wherever necessary. For the cost benefit evaluation of the system, we fix the various costs as $c_1$=15, $c_2$=20, $c_3$=12, $c_4$=25, and $c_5$=8.

Figure 3 depicts the effect of $\lambda$ on expected system size when the main server is in busy period of FES (or) SOS ($E[L_b]$) for various SOS service rates of main server ($\mu_2$). We can see from the graph that for any $\mu_2$, $E[L_b]$ expectedly increases with the increase of $\lambda$. Furthermore, for fixed $\lambda$, opposite effect is observed with increase of $\mu_2$ because of faster services.

Figure 4 presents the impact of arrival rate $\lambda$ on expected system size when standby server is working ($E[L_s]$) for different values of service rates of standby server ($\mu_3$). The graph shows that as $\lambda$ rises, $E[L_s]$ rises as well with fixed service rate $\mu_3$. Further, for fixed $\lambda$, increase of $\mu_3$ results in decrease of $E[L_s]$, which is indeed true. Figure 5 shows

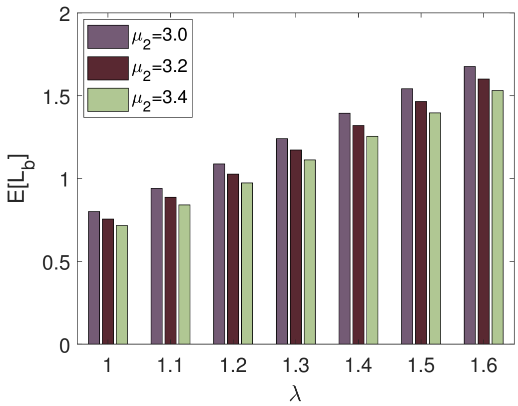


FIGURE 3. Impact of $\lambda$ on $E[L_b]$ for various $\mu_2$

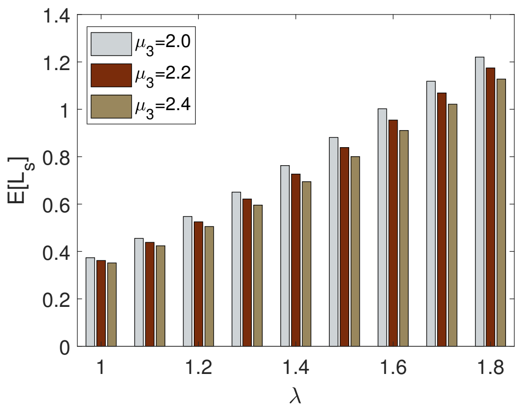


FIGURE 4. Impact of $\lambda$ on $E[L_s]$ for various $\mu_3$

impact of the main server's service rate during SOS ($\mu_2$) on $E[L_b]$ for different values of $\beta_2$. For fixed $\beta_2$, it can be seen that $E[L_b]$ decreases, as expected as $\mu_2$ rises.

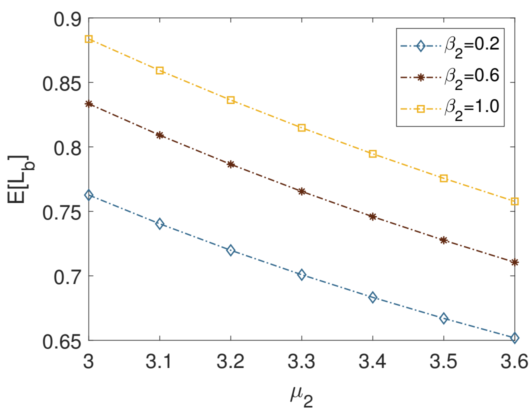


FIGURE 5. Impact of $\mu_2$ on $E[L_b]$ for various $\beta_2$

Figure 6 illustrates the influence of repair rate ($\gamma$) on the anticipated system size when main server is broken down and the standby server is not operational ($L_{smb}$). We can observe from the graph that as $\gamma$ increases, $L_{smb}$ decreases. This is due to the fact that increase in the repair rate has a positive effect on the availability of the main server for resuming his service, thereby decreases the system size and waiting time. For fixed $\gamma$, the system size increases due to the higher threshold value $q$.

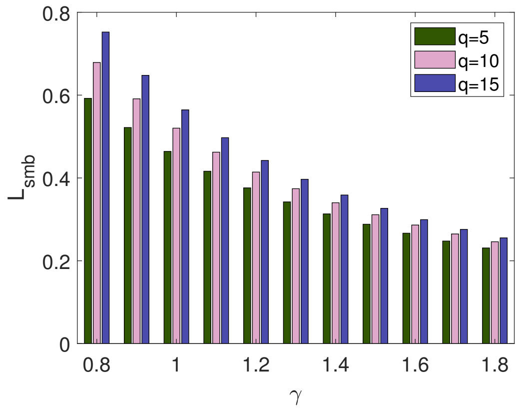


FIGURE 6. Impact of $\gamma$ on $L_{smb}$ for various $q$

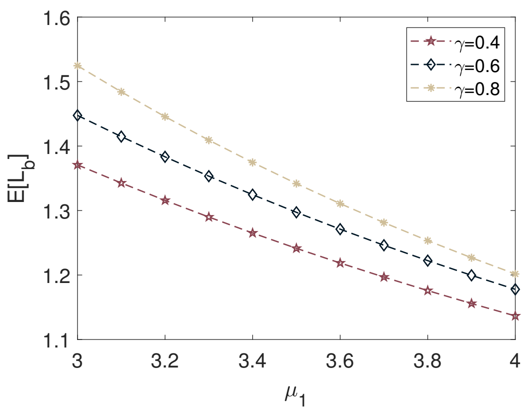


FIGURE 7. Impact of $\mu_1$ on $E[L_b]$ for various $\gamma$ values

Figure 7 presents the effect of service rate of main server during $FES(\mu_1)$ on $E[L_b]$ for various values of repair rate($\gamma$). It can be observed that for fixed $\gamma$, $E[L_b]$ drops as expected as $\mu_1$ increases. Additionally, with fixed $\mu_1$, a rise in $\gamma$ increases the presence of the primary server, resulting in a decrease in $E[L_b]$

Figure 8 illustrates the effect of the service rate $\mu_1$ on the total cost function $f(\mu_1)$ for three different values of $r$. For each value of $r$, the cost initially decreases as the service rate $\mu_1$ increases, reaching a minimum, and then starts to rise again as $\mu_1$ continues to increase. This behavior is typical in queueing systems, where there is an optimal service rate that minimizes total costs, balancing the trade-off between low and high service rates. For $r = 0.6$, the system achieves the lowest minimum cost at a relatively lower service rate. This suggests that with a smaller $r$, the system operates more efficiently, allowing for cost minimization at a lower service rate. As $r$ increases, both the minimum cost and the optimal service rate rise, indicating that higher values of $r$ make it more difficult to minimize costs, thus requiring a higher service rate to achieve optimal performance.

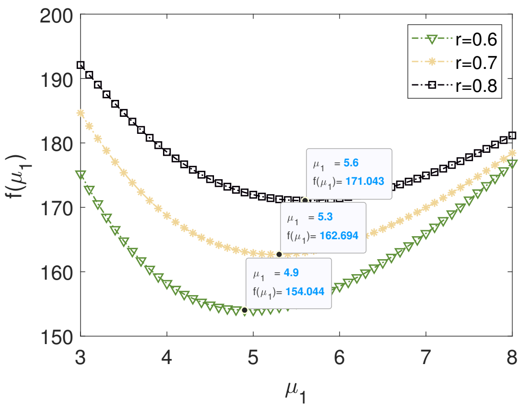


FIGURE 8. Impact of $\mu_1$ on $f(\mu_1)$ for various $r$ values

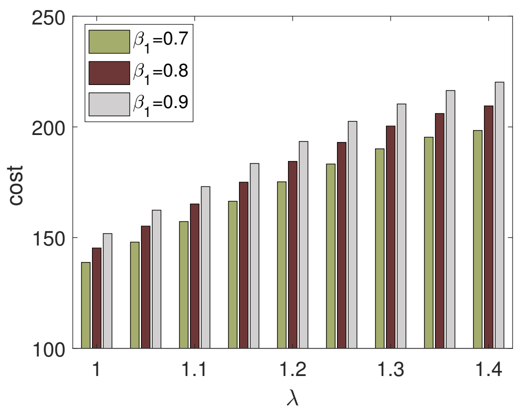


FIGURE 9. Impact of $\lambda$ on cost for various $\beta_1$ values

Figure 9 demonstrates the relationship between the arrival rate $\lambda$ and the cost function for different breakdown rate values $\beta_1$. As the arrival rate $\lambda$ increases, the cost function consistently increases, regardless of the breakdown rate. This trend suggests that higher arrival rates lead to increased costs due to higher system utilization and the increased likelihood of system breakdowns or inefficiencies. The graph shows this behavior across three different values of $\beta_1$, where a higher breakdown rate generally results in a higher cost for a given arrival rate. Thus, both the arrival rate and the breakdown rate significantly influence the overall cost of the system.

## 7. CONCLUSIONS

This study focused on a Markovian queueing system that has cold standby servers, SOS, server failures. We derived important theoretical and practical insights. Below, we summarize the key results, practical implications and potential directions for future research:

- Key results and practical implications: The study employed a matrix geometric approach to solve the solution, taking advantage of its block structure to obtain steady state probabilities and performance measures. The stability condition and the cost functions were derived. Graphs were depicted to analyze the influence of various factors on the performance of the system.
- Future work: Potential extension of our model could more complex queueing scenarios, such as queue-dependent multi-server systems, differentiated vacation policies for servers, inclusion of customer impatience behaviors such as reneging, balking, and jockeying, as these phenomena can significantly affect system performance and customer satisfaction.

## References


[1] Madan, K. C., (2000), An $M/G/1$ queue with second optional service, Queueing Syst., 4(1), pp. 37-46.

[2] Jain, M. and Chauhan, D., (2012), Working vacation queue with second optional service and unreliable server, Int. J. Eng., 25(3), pp. 223-230.

[3] Vijaya Laxmi, P., Bhavani, E. G. and Kumar, R., (2020), Correlated reneging in an optional service Markovian queue with working vacations, Reliability: Theory & Applications, 15(4), pp. 102-116.

[4] Vijaya Laxmi, P., Qrewi, H. A. and George, A. A., (2022), Analysis of Markovian batch service queue with feedback and second optional service, Reliability: Theory & Applications, 17(2(68)), pp. 507-518.

[5] Vijaya Laxmi, P., Bhavani, E. G. and Jyothsna, K., (2022), Analysis of Markovian queueing system with second optional service operating under the triadic policy, OPSEARCH, 60(1), pp. 256-275.

[6] Ke, J. C., Chang, C. J. and Chang, F. M., (2010), Controlling arrivals for a Markovian queueing system with a second optional service, Int. J. Ind. Eng., 17(1), pp. 48-57.

[7] Rozi, A. and Gupur, G., (2013), Another eigenvalue of the $M/M/1$ queueing model with optional second service, J. Pseudo-Dif. Oper. Appl., 4, pp. 413-441.

[8] Gray, W. J., Wang, P. P. and Scott, M., (2000), A vacation queueing model with service breakdowns, Appl. Math. Model., 24(5-6), pp. 391-400.

[9] Tarabia, A. M., (2011), Transient and steady state analysis of an $M/M/1$ queue with balking, catastrophes, server failures and repairs, J. Ind. Manage. Optim., 7, pp. 811-823.

[10] Seenivasan, M., Manikandan, H. and Epciya, J. S., (2022), $M/M/1$ Queue with server breakdown, single working vacation, state reliant customers and feedback, In 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, pp. 1-5.

[11] Rao, S. H., Kumar, V. V., Kumar, B. S. and Rao, T. S., (2021), Analysis of two-phase queueing system with impatient customers, server breakdowns and delayed repair, Int. J. Pure Appl. Math., 115(4), pp. 651-663.

[12] Chakravarthy, S. R. and Kulshrestha, R., (2020), A queueing model with server breakdowns, repairs, vacations, and backup server, Oper. Res. Perspect, 7, pp. 100131.

[13] Braglia, M., (1993), $M/M/1$ queuing model with ordinary maintenance and breakdowns, RAIRO Oper. Res., 27(1), pp. 1-21.

[14] Gupta, P. and Kumar, N., (2021), Performance analysis of retrial queueing model with working vacation, interruption, waiting server, breakdown and repair, J. Sci. Res., 13(3), pp. 833-844.

[15] Saritha, Y., Devi, V. R. and Chandan, K., (2020), $M/M/1$ queue with breakdowns, two varieties of repair facilities, timeout and vacation, J. Mech. Cont. & Math. Sci., 7, pp. 215-224.

[16] Khalaf, R. F., Madan, K. C. and Lukas, C. A., (2011), On a batch arrival queuing system equipped with a stand by server during vacation periods or the repairs times of the main server, J. Probab. Stat., 22(1).

[17] Ayyappan, G. and Karpagam, S., (2019), Analysis of a bulk queue with unreliable server, immediate feedback, N-policy, Bernoulli schedule multiple vacation and stand by server, Ain Shams Engineering Journal, 10(4), pp. 873-880.

[18] Ayyappan, G. and Karpagam, S., (2020), Analysis of a bulk service queue with unreliable server, multiple vacation, overloading and stand-by server, Int. J. Math. Oper. Res., 16(3), pp. 291-315.

[19] Vijaya Laxmi, P. and Bhavani, E. G., (2021), A repairable second optional service queueing system with warm and cold standbys, International Journal of Reliability and Safety, 15(1-2), pp. 104-122.

[20] El-Rayes, A., Kwiatkowska, M. and Norman, G., (1999), Solving infinite stochastic process algebra models through matrix geometric methods, School of Computer Science Research Reports-University of Birmingham CSR.

[21] Lakshmi, K. and Ramanath, K., (2014), A matrix geometric approach to the $M/M/1$ two-phase multi optional retrial queue with Bernoulli feedback, impatient customers and a server subject to breakdown and repair, Opsearch, 51, pp. 36-49.

[22] Joshi, P. K., Gupta, S. and Rajeshwari, K., (2021), Matrix geometric method for the analysis of $M/M/1$ model under repair, Adv. Appl. Math. Sci., Volume 20, Issue 7.

[23] Shah, W., Shah, S. A. A., Soomro, S., Khan, F. Z. and Menghwar, G. D., (2009), Performance evaluation of multisatge service system using matrix geometric method, In 2009 Fourth International Conference on Systems and Networks Communications, pp. 265-269.

[24] Neuts, M. F., (1981), Matrix geometric solutions in stochastic models, volume 2 of Johns Hopkins Series in the Mathematical Sciences.

**Dr. Pikkala Vijaya Laxmi** is a Professor in the Department of Applied Mathematics at Andhra University, Visakhapatnam, India. She received her MSc and PhD degrees from Indian Institute of Technology, Kharagpur, India. Her research interests include continuous and discrete-time queueing models and their applications. She has publications in various journals like *Operations Research Letters, Queuing Systems, Applied Mathematical Modelling, International Journal of Applied Decision Sciences, Journal of Industrial and Management Optimisation (JIMO)*, etc.

**Gilaka Anjalidevi** is a holder of the degree of Bachelor of Science in Mathematics from the Andhra University, Visakhapatnam. She received her Master of Science in Mathematics from Andhra University, Visakhapatnam. Currently, she is a PhD scholar in the Department of Applied Mathematics at Andhra University, Visakhapatnam. Her research interest is queueing theory and its applications.